The Dissertation Committee for Rebekah Elizabeth Post certifies that this is the approved version of the following dissertation:

The impact of social factors on the use of Arabic-French code-switching in speech and IM in Morocco

Committee:

Barbara E. Bullock, Supervisor

Carl Blyth

Alexandra K. Wettlaufer

Kristen Brustad

Almeida Jacqueline Toribio

The impact of social factors on the use of Arabic-French code-switching in speech and IM in Morocco

by

Rebekah Elizabeth Post, B.A., M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2015

Acknowledgements

The first thanks go to my amazing committee members who gave me never ending advice in addition to countless suggestions and edits to help keep things moving and on the right track. To my supervisor Barbara Bullock, I don't know what I would have focused on if you hadn't arrived here during my second year and there is no way to begin to acknowledge all that you have done. All of your insight and advice has been invaluable. To my mentor on the Arabic side, Kristen Brustad, thank you for introducing yourself to me in 2011 when I had started down this path but really had no idea what I was getting into. Thank you for encouraging me to learn Arabic and then making sure it happened, for being so willing to share your time and experience, and for including me in departmental activities. To Carl Blyth, thanks for all of your support in my varied interests from my first year through dissertation work with some much-needed reality checks at key points along the way. To Jacqueline Toribio for your untiring support, care and editing as you do everything you can to help her students succeed as scholars and as individuals. And to Alex Wettlaufer for joining us later in the game and providing a reminder to write for those from all backgrounds who might be interested in language contact in Morocco.

Many of my fellow students have also been important in finishing this work. Thank you to all of the CADRiin (Center for Arabic Dialects Research) for your acceptance, advice, and patience with my questions about Arabic, and to those in the Moroccan Table who help me to stay connected to the language and excited about my research. On the other side, many thanks to the FRIT grad students, past and present, who shared the joys and difficulties of grad school and who I got to know over coffee, study sessions, and happy hours as we commiserated and celebrated the ups and downs.

ألف شكر لكل المغاربة للي هضروا معيا و جلست معهم ساعة أو شحرين – عائلتي المغربية خديجة وحسان, ختي كريمة للي شجعتني في ميريكان وفي المغرب وكل الشباب هضرنا مع بعض على البحث أو الحياة. انتم ديمة في أفكاري.

The monumental task of transcription was made possible only through the help of many others; Aicha Moujane who sat with me explaining Moroccan language and cultural references for more hours than I can remember, Majda Zouak who made it possible to continue the work in the US, Dr. Ahmed Echcharfi who helped to finish and correct the transcriptions, and Mike Turner who transcribed a conversation and freely shared his many views on Moroccan lexicon, grammar, and phonology. Des autres m'ont aidée dans le recrutement des participants – mes belles Belen et Firdaws qui m'ont tellement aidée avec tout ce qu'elles pouvaient faire, ainsi que les femmes de l'équipe TANIT à l'Université de Moulay Ismail.

Perhaps the biggest debt goes to my family, biological and chosen, for their support, encouragement, inspiration, love, and reminders of the world outside of studies. For laughing at books, movies and ourselves, cooking for 2 or 100, relaxing (or was it procrastinating?), singing karaoke, working, running, swimming, complaining, exploring new cities and countries, discussing life, love and politics, dancing in the kitchen, club, and living room, writing papers day and night, and partying together – for always being there for me.

This research project was supported by the National Science Foundation (NSF) doctoral dissertation improvement grant #1322320. Any findings, opinions, and conclusions expressed in this dissertation are those of the author and do not necessarily reflect the views of the National Science Foundation.

The impact of social factors on the use of Arabic-French code-switching in speech and IM in Morocco

Rebekah Elizabeth Post, PhD

The University of Texas at Austin, 2015

Supervisor: Barbara Bullock

The use of French in code-switching (CS) with Moroccan Colloquial Arabic (MCA) has been explored qualitatively in a number of studies, but quantitative methods have rarely been applied to CS in this language pair. Research on CS patterns as a function of extra-linguistic factors has similarly received little attention, despite the implication in many studies that these factors are significant in the use of CS. This dissertation seeks to address these gaps in the literature by quantitatively examining the use of Arabic-French CS by young adult speakers of MCA in spoken and written information communication.

This study examines three extra-linguistic factors in speech and Instant Messaging (IM): Sex, French Proficiency, and Language Attitude. The analysis reveals that male speakers are significantly more French in written IM. Positive attitude toward French and MCA-French CS has a highly significant impact on the rate of French employed in spoken conversation. Meaningful results are also found for the French constituents employed in CS with regard to each of the extra-linguistic factors. Notable differences are found between sexes in the types of French constituents used in both communication modes, as well as for speakers of different French proficiency levels.

The categorization of French-origin nouns as instances of CS or borrowing is also explored by considering multiple aspects of use of these lexical items. A number of French-origin nouns, absent from dictionaries of MCA, are proposed to now be borrowed into the dialect. The analysis also reveals a number of French-origin words that are used by a number of speakers, but remain instances of CS.

The results of this investigation highlight the importance of quantification in studies of CS and provide data for comparison with other corpora from this and other language pairs. The differences identified in CS by communication mode indicate that there is a need for a model of written CS that accounts for the unique characteristics of this mode. Finally, little work has been published on the relationship between extra-linguistic factors and structural patterns in CS, but the current results suggest that the impact of social factors should not be ignored when considering structural aspects of CS.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

The following abbreviations are used in the syntactic glosses of examples throughout the dissertation

1st	first person (singular when not overtly marked for plural)
2nd	second person (singular when not overtly marked for plural)
3rd	third person (singular when not overtly marked for plural)
DEM	demonstrative determiner
EMPH	emphasis
F	feminine
FUT	future
M	masculine
NEG	negation
PL	plural
PRES	Present tense (in Arabic, more accurately مضارع)
SG	singular

1. Introduction

Code-switching (CS), or the use of two languages within a single utterance, is a common practice across many bilingual speech communities. However, not all bilinguals code-switch and the observed level of CS within a community, and attitudes toward it, vary greatly. Far from the popular characterization of CS practices as random or evidence of low proficiency, CS displays consistent syntactic patterns, although these patterns may vary by community (see Muysken 2000 for an overview of these differences). What leads some speakers to strictly separate their languages while others may switch between languages multiple times in a single sentence? This dissertation seeks to address this question through analyzing the informal spoken and written production of Arabic-French CS by 36 young adult native speakers of Moroccan Colloquial Arabic.

CS has often been studied in the context of French-Arabic contact in Morocco (Abassi 1977, Bentahila 1983, Bentahila and Davies 1983, 1995, Ziamari 2008, among others), in Tunisia (Stevens 1974, Belazi 1991, Lawson & Sachdev 2000, Sayahi 2011a, among others), and in the Moroccan diaspora in Quebec (Nait M'Barek and Sankoff 1988, Redouane 2005). Despite the common occurrence of this phenomenon, popular opinion of it is often negative (Bentahila 1983, Lawson and Sachdev 2000). Early research on code-switching concurred with this point of view, suggesting that bilingual speakers' alternation between languages constitute a random, unstructured mixture (Weinrich 1953) but it has since been consistently shown that identifiable grammatical patterns exist in CS (Timm 1975 as the first of many) wherever it occurs and that native speakers who participate in CS can provide reliable grammaticality judgments on acceptable switching types (Anderson 2006, Post 2010, Grabowski 2011).

CS between Arabic and French is often noted by lay speakers as well as language researchers. In discussions of dialects with Arabic speakers, one recurring stereotype of North African dialects, and Moroccan in particular, is that the dialect is 'corrupted' by French. Moroccan Colloquial Arabic (hereafter MCA or Moroccan Arabic) is stereotyped as being heavily influenced by French, containing many borrowed words and punctuated by frequent code-switching with French (Hachimi 2013). This stereotype is echoed by students learning Arabic who voice concern that the Moroccan dialect is harder to learn because of its use of French lexical items. Examples drawn from Bentahila and Davies (1983) may seem to give evidence of this, as in (1.1) where the speaker began his/her utterance in MCA but finished it using a French relative clause.

(1.1) taybɣiw yšufu ši ħaža *qui est différente*
PRES.3rd.like.PL 3RD.see.PL some thing *that is different.F*
"They enjoy looking at something that is different"
(=13 in Bentahila & Davies 1983)

It is undeniable that French lexical items have been borrowed into the Arabic dialects of North Africa, as English lexical items have similarly entered many other Arabic dialects. However, no quantitative studies are available on the proportion of French used in CS with MCA that would indicate whether the impressionistic judgment of MCA speakers' high use of French is accurate or whether the role of French in MCA today is, instead, a modest one.

1.1 AIMS OF THIS DISSERTATION

In this dissertation I explore the use of French-origin lexical items by Moroccan young adults in MCA-French code-switching in order to understand the amount of French lexical items they use in CS with MCA and the structure of these MCA-French utterances. For this reason, this research has two equally important goals: first, to quantify the use of

French-origin lexical items in order to understand whether this is truly a strong tendency that marks this dialect as distinct from that of other countries and regions. The second is to explore the relationship between extra-linguistic factors and the use of French in MCA-French code-switching in order to understand how the use of French lexical items varies among current speakers.

In order to reach these goals, the following research questions will be explored throughout this dissertation:

Question 1. What proportion of informal language use by young Moroccans is constituted by French lexical items?

Question 1b. What is the structure of the French lexical items in MCA?

Question 2. Does communication modality (written or spoken) affect the rate or structure of French in MCA-French CS?

Question 3. How do external factors of SEX, FRENCH PROFICIENCY or LANGUAGE ATTITUDES contribute to the variation observed in rates and structure of CS?[1]

These questions are investigated using a corpus of informal communication collected in the fall of 2013. The corpus contains data from two informal production tasks carried out in same-sex speaker dyads: a spoken in-person conversation and a written instant messaging (IM) chat. The use of two separate tasks with the same dyads creates parallel corpora in which it is possible to analyze how communication modality impacts CS. Background data was also collected from these speakers in order to obtain information regarding the extra linguistic factors under study.

The unique methodology of this study allows us to comment on several aspects of language use. The use of two communication modalities responds in part to Gullberg,

[1] These terms are given in small caps when used as extra-linguistic factors and in standard type face when used to refer to characteristics of a given speaker or speakers.

Indefrey and Muysken's (2009) recommendation to use multiple methods to access speaker knowledge to better understand CS practices; the analysis of both rate of each language in CS and of constituents used in French, as opposed to Arabic, represents another way to combine methods in the study of this phenomenon. I take a quantitative approach to the use of CS in order to identify the actual impact of French in MCA-French CS by examining the rate of French lexical items and the constituents in which French lexical items appear. The quantification of rate and structure of French lexical items in MCA-French CS then allows observation of differences between the written and spoken modalities. Following overall quantification by modality, the target extra-linguistic factors can be examined in spoken and written communication.

1.1.1 Role of quantification in linguistic studies of CS

Quantification is increasingly important in many subfields of Linguistics, but it is typically absent from studies of CS. Rate of CS is almost never reported, perhaps due the rare nature of the phenomenon; when information about rate of CS is given, it is often done indirectly by reporting the number of switches in the full corpus. The data that forms the base of Poplack's (1980) highly influential study contains a total of 1,835 switches in sixty-six hours of data, or just 27.8 switches per hour. In contrast, Sayahi's (2011a) corpus includes an immensely higher number of switches, totaling 1,721 in his three hour corpus (or 573.7 switches per hour). The spoken corpus collected for the present study contains 1,368 French constituents in 4.25 hours of speech (321.9 per hour), or about 250 switches per hour fewer than in Sayahi's data. Awareness of the true level of occurrence of CS in the target communities can indicate whether it is a common practice, or peripheral to typical communication and notable primarily for its salience. While the use of French by Arabic speakers in the latter studies is higher than that of English among the Spanish

participants interviewed for Poplack's study, calculating the rate of switches per hour can be deceiving because the number of switches per hour does not take speech rate into account. The participants in the current study employ a median average of 2.6% French lexical items in their speech.

The type of constituents switched are often reported only anecdotally with a strong focus on viewing constituent structure categorically. That is, researchers often only report whether a syntactic category is observed as an other language embedding or not. However, reports of CS that quantify constituents indicate that the structures used in CS are potentially socially meaningful. Poplack (1980) found the distinction between intra- and inter-sentential constituents to be an indicator of proficiency; Sayahi's data (2011a) reveals that less educated male speakers use a higher proportion of French bare nouns than their university-educated counterparts; Bentahila and Davies (1995) observed that those more proficient in French employ a greater proportion of full sentences in French. Such findings suggest the importance of the structure of CS in understanding differences between speaker groups.

Quantification is considered of utmost importance in the current study because it provides a straightforward way to compare between speaker groups and data sets. Careful quantification of the patterns found in CS can indicate those patterns that are most common and widespread while marking others as marginal. Consideration of how such patterns are used, and by which speakers, can shed light on their social significance among target speakers. By contrast, qualitative study of CS is often categorical in that studies tend to focus on whether a pattern is present in a given corpus or absent from it. A strictly qualitative analysis cannot indicate the strength of a trend within a given data set or reveal the level of variation found within or between speakers. By nature, qualitative reports include limited examples of target phenomena. Relying solely on qualitative results can be

misleading in that researchers may read the same isolated examples and draw opposing conclusions from them.

One instance of drawing conclusions from isolated examples can be found in Myers-Scotton's (2008) discussion of data in Ziamari (2003). Ziamari provides an example of MCA-French CS that violates the System Morpheme Principle, a key principle of Myers-Scotton's Matrix Language Frame model. Myers-Scotton agrees that the example presents counter-evidence to the System Morpheme Principle, but dismisses the difficulty raised by Ziamari's example by stating, "but remember that it is only a single example" (32). The accuracy of the MLF model is not taken up here; my concern lies in the line of reasoning. Only careful quantification of the frequency of observed patterns of CS can indicate whether a 'single example' is an outlier (i.e., speech error, an anomaly) that can safely be discarded, or a common pattern within the speech community that might indeed call into question the System Morpheme Principle on a broader level.

Another common type of dismissal of data occurs when there is disagreement between native speakers with regard to whether the example appears to be authentic. MacSwan (1997) states that "some of the examples of Spanish-English code switching in Belazi, Rubin and Toribio (1994) are spurious" (186) because his own informants did not find them acceptable, this despite the fact that one of the authors of that study is a code-switching Spanish–English bilingual. His stance ignores the possibility that different speakers from distinct dialects or communities may use the same language pair in divergent ways. For instance, the judgment data that I collected from Moroccan and Tunisian Arabic speakers (Post 2010) reveals that speakers from each country report significantly different acceptability of Arabic-French code-switches that occur within the Determiner Phrase. Thus, quantification of naturalistic data allows researchers to compare the strength of specific patterns present in their corpora with the patterns found in other data sets. Those

working on grammatical models of CS, such as Myers-Scotton and MacSwan, will then have more robust data on which to base their own proposals. Quantification, and a sound interpretation of it, is critical to understanding the probability of occurrence of patterns in language data.

At the time of the earliest studies on code-switching, quantification was a protracted process that required manual identification of target words and structures. With a focus on grammatical constraints that were potentially revelatory of universal properties of syntax, it is understandable that these studies were largely qualitative. But today, Natural Language Processing (NLP) methods allow for the rapid extraction of target lexical items and for the automatic identification of linguistic structure for many languages. Yet the variability of Arabic dialects and the lack of description of them presents a difficulty for achieving a fully automatic NLP analysis of colloquial Arabic, (Diab et al. 2010, Almeman and Lee 2013, Cotterell et al. 2014) and data involving code-switching with an Arabic corpus presents even greater challenges to automatic processing of a corpus because the system must be trained to recognize two potential lexicons and grammatical systems. In spite of these challenges, this study combines the traditional (time-consuming) methods of manual identification of language identification for each lexical item with an automated NLP analysis to provide a detailed representation of the use of French in MCA-French CS.

1.1.2 Comparison between modalities

In this dissertation I explore the use of CS in Morocco as naturally produced in spoken and written informal conversations. Past research in CS that focuses on this language pair has been limited to speech, the traditional context of CS in Morocco due to the informal nature of CS and the formal nature of most written genres (see §2.2 and

Chapter 4). However, modern technology has created an environment in which the dialect is now utilized in written form. The inclusion of written CS thus provides a first step in understanding the practices of writing Moroccan Arabic, the language spoken by 89.8% of Moroccans (Haut Commissariat du Plan 2004). This dialect, like other Arabic dialects, is not typically written, and the basic sound-symbol correspondences noted in §4.3.4 are a first step in documenting the written form of MCA.

Consideration of both spoken and written interaction provides two distinct modalities in which to analyze CS. Even in semi-synchronous chat environments, speakers must make orthographic choices that can reflect morphosyntactic features due to the higher level of planning required in written communication. These choices may include which words to use in one language versus the other or which type of article to use on nouns inserted from French into Arabic (see Chapter 7). Similarities between these modalities may indicate conventionalized patterns within the community, whereas divergent usage in each modality may suggest areas where variation remains or where a conventionalized pattern is restricted to only one of the modalities.

The analysis of spoken and written CS of the same speakers is unprecedented in published research to my knowledge. Differences identified in monolingual spoken and written language (Biber 1988) and monolingual spoken and informal written Computer-Mediated Communication (CMC) specifically (Crystal 2001, 2006, Tagliamonte and Denis 2008) indicate that important differences exist between communication modes, and that written CMC is a new modality that reflects structural and lexical properties of both traditional modalities, combined in a unique way that require us to analyze it as a separate modality in its own right. The comparison of CS in informal speech and written IM therefore provides evidence of differences between these modalities as used by MCA/French bilinguals and another way to access grammatical aspects of CS.

1.1.3 Impact of social factors on the use of CS

Although a basic level of bilingualism is required to participate in CS, a great amount of variation is observed between language pairs and between communities where the same language pair is used. These differences include whether CS is acceptable at all, reflecting language attitudes, who participates in it, what functions it serves, and how the languages pattern morphosyntactically when present in a single utterance. Competence in each language seems to affect the use of CS, but whether this has the greatest impact at the level of clause (Poplack 1980) or at the level of the constituent (Bentahila & Davies 1995, Sayahi 2011a) appears to vary by community. There are doubtlessly social factors at play, but the extent to which macro-social variables such as gender contribute to the variation observed in code-switching is unclear in certain contexts (Ibrahim 1986, Wodak & Benke 1997, Chakrani 2010, Sayahi 2011a).

After reviewing extralinguistic factors that have been suggested to be relevant in Arabic-French CS, SEX, LANGUAGE ATTITUDES, and PROFICIENCY were each examined in reference to the rate of French used by participants and the constituents in which the French lexical items appear. One potential factor, AGE, is kept constant by restricting the age range to young adults at or near university age. This age group was selected due to the typical role of young adults in language change, as they are often the first to adopt new forms, and thus may indicate the direction of linguistic change within a community (Labov 2001). As I will demonstrate, the analysis indicates that each of the three factors impacts CS in different ways, as seen in Chapter 6.

1.2 MAJOR CONTRIBUTIONS OF THIS DISSERTATION

The innovative methodology and the resulting corpus collected for this dissertation results in contributions to the fields of CS, bilingualism, language modality, writing

systems, CMC and Arabic dialectology. Five major contributions of this dissertation and its findings are discussed in this section. Additional contributions are possible through further analysis of the corpus from which the data is drawn.

The corpus created for this research is itself a major contribution to Arabic dialectology. The 4.5 hours of spoken conversation transcribed for this analysis represent over 40,000 words. The current transcription represents fifteen minutes of conversation from each participant dyad; if participants spoke at the same rate throughout their hour of conversation, transcription of the spoken corpus will total approximately 160,000 words. This may seem trivial when compared to the spoken portion of the British National Corpus (BNC Consortium 2007), which consists of 10 million words, or the Corpus of Contemporary American English (Davies 2008-) with 90 million words from speech. However, at the time of writing only approximately 80 minutes of transcribed spoken Moroccan Arabic is freely available. This resource is provided by CorpAfroAs, with 45 additional minutes scheduled to be added (Mettouchi, Vanhove & Caubet 2012). In light of the dearth of resources on MCA, the current corpus will more than triple the transcribed spoken data freely available in this dialect.

Another aspect of the contribution to Arabic dialect studies and CS is the quantification of French lexical items used. In contrast to stereotypes, the results presented in the following chapters reveal a low level of French lexical items used in Moroccan Arabic by the participants. The median spoken rate of French lexical items is only 2.6% with a mean of 5.2% due to a small number of higher CS users. The fifteen minutes extracted for transcription and analysis in this dissertation were selected to include the topic of education as this was present in all of the conversations and impressionistically expected to contain a higher rate of French lexical items than other topics. Thus, the average rate of French lexical items found in speech is expected to decrease when the full conversations

are transcribed. The quantification of CS among young Moroccans also serves as an update to the data reported by Bentahila and Davies (1995), which was collected an estimated 30 years before the current corpus and which, until now, constituted the only published quantification of MCA-French CS in Morocco.

The comparison of CS in two modalities is yet another contribution of this work. The rate of French is vastly different in each modality: the mean average of French lexical items in written IM is 9.6%, and falls to 5.2% in spoken conversation. It may be that the written conventions include a greater amount of French, at least in IM through the Facebook website. The types of French constituents used in each modality is also substantially different; the written IM data contains many more extrasentential elements, as seen in (1.2) below, while nouns are more common in spoken, as in (1.3).

(1.2) *oui* omachi fdar db
yes and.not in.house now
"*yes*, and I'm not at home now" (P11225, written)[2]

(1.3) kanħli *l'appareil* ħta katsxən
PRES.1st.leave *the=appliance* until PRES.3rd.F.heat
"I leave *the appliance* until it gets hot" (P11021)

The notable difference in rate of French lexical items and the divergences in types of French constituent employed in each modality suggests differences between the two that may be related to discourse structure and the role of planning. While no studies can be found on CS practices of the same speakers in these two modalities, work on monolingual English points to a variety of differences that justify such a methodology (Tagliamonte & Denis 2008).

[2] The use of French lexical items is indicated by the use of *italics* in all examples. This text, and all examples drawn from the written corpus, is left in the orthography as produced by the participants. See §4.3.4 for more on the emerging written system of MCA.

The analysis of extra-linguistics factors reveals meaningful distinctions in French use between the participants, some of which are clear across both spoken and written modalities, while others can be observed only in one. SEX is found to impact the rate of French only in writing, while the constituents used by members of each sex show certain similarities by sex and others by modality (§6.2). Among the present speaker group, FRENCH PROFICIENCY has surprising results across modalities; it has no effect on rate of French, and few generalizations can be made for this factor by constituents used (§6.3). LANGUAGE ATTITUDE also shows sensitivity to modality as the rate of French displays a significant difference in spoken, but not written, conversation. However, in both modalities speakers with a more positive attitude toward French use a broader range of French constituents in their production (§6.4). The impact of SEX and LANGUAGE ATTITUDE on the use of French in MCA-French CS indicate that significant differences in CS practices exist within communities. It cannot be assumed that a group of speakers from a single community will use the same rate or the same syntactic structures in CS; extra-linguistic factors must be taken into account to establish an accurate description of CS practices.

Finally, this study adds to the conversation regarding the distinction between borrowing and code-switching. It is clear in the current corpus that relying on dictionary inclusion as an indicator of lexical borrowing is grossly inadequate in this context. This methodology may be sufficient for well-documented languages with a lexicographic tradition, but this is not the case for MCA, Arabic dialects, or many other less prestigious language varieties. Poplack and colleagues (Poplack, Sankoff & Miller 1988, Poplack, Zentz & Dion 2011, and others) state that morphosyntax should instead to be used to determine whether a word has been borrowed or remains an instance of CS, but the current results indicate that this distinction is difficult to make when morphosyntactic elements may

be borrowed along with the lexical item in question. A solution to this difficulty remains for future work; I hope that the current data will be useful in that endeavor.

1.3 ORGANIZATION

In order to understand how languages are used in a single country, we first must understand the languages that are involved and how they came to be found in the setting under study; this is undertaken in Chapter 2. Today six language varieties are typically included in discussions of language contact in Morocco: Modern Standard Arabic, Moroccan Colloquial Arabic, Tamazight, French, Spanish, and English. Moroccan Arabic and Tamazight (more accurately, the three primary Tamazight dialects) are the native languages of Morocco. Modern Standard Arabic (hereafter SA) is the primary language of education and one official language of the country; the other official language is Tamazight. The history and roles of MCA and French are also described in this chapter, despite their lack of official status.

Chapter 3 adds to the background information with a discussion of code-switching including the social factors that may impact the use of CS. This chapter also considers past findings on Arabic-French CS because the language pair is often addressed in research on CS as an example of mixing between two languages that are typologically distinct.

In Chapter 4, I turn to a new environment for CS: the internet. CS is most often examined as a spoken phenomenon. This is the typical domain of its use, as CS is generally found in informal communication, which has traditionally been spoken. However, widespread access to the internet and the rise of informal written communication through it has provided an opportunity for speakers to use informal language practices in writing. [3]

[3] The term 'speaker' is used as the default term in this work whether the communication produced is spoken or written as individuals included in cited research who write in the target varieties also speak it.

MCA is not the only 'unwritten' language to be found in this mode; for that reason, Chapter 4 explores studies of CS on the internet in a variety of language pairings with emphasis on those that involve one traditionally unwritten variety.

Based on the findings from the previous chapters, the methodology of the current study is presented in Chapter 5. Mixed methods are employed in order to provide a more complete picture of language use among the target speakers. The corpus for the current study was gathered via two production tasks: one spoken, face-to-face conversation and one written IM conversation between the same speaker pairs. The production tasks are complemented by ethnographic information gathered via a written survey. A short French proficiency test was also administered in order to identify whether this factor may play a roll in CS practices. In addition to detailing the rationale for the research design given above, this chapter also addresses the transcription process and coding, both through Natural Language Processing by computer and manually coding the data.

The results of the study are given in Chapter 6 beginning with the rates and patterns of CS found in each modality. I then focus on the three extra-linguistic factors of SEX, FRENCH PROFICIENCY, and LANGUAGE ATTITUDE in order to explore whether these factors affect the rate or structure of MCA-French CS as produced by participants. The potential impact of these factors is examined using logistic regression in order to identify the factors that are statistically significant in the rate of French in MCA-French CS. This chapter also notes important findings that are similar to those of past studies as well as the results that are unique to the current data and merit further investigation.

Chapter 7 contains a secondary inquiry based on questions raised during the process of transcription and analysis. Reliance on dictionaries to identify borrowings may be inadequate when working with an under-documented variety such as MCA. I noticed many repeated French nouns during the transcription process, including many for which MCA

translation equivalents were not present in the corpus, leading to further exploration of this topic. The most commonly used nouns were identified, providing thirty French lemmas, and then analyzed using four metrics in order to evaluate their status as borrowings into Moroccan Arabic or code-switches that remain primarily French. The analysis suggests that certain lexical items may be stable loanwords in MCA while others likely remain code-switches. These results are preliminary because of the limited nature of the corpus used to analyze these nouns and the restricted speaker group included within it.

The dissertation concludes in Chapter 8 with a summary of the findings and a discussion of their implications across related fields of study. Many future directions are also given due to the fact that no results on a living language can be absolute; speakers continue to speak and write in MCA. The methods and results of this dissertation aim to encourage similar analysis of this dialect and other contact varieties in order to improve comparability between data sets and their results.

1.4 A NOTE ON THE TRANSCRIPTION SYSTEM

The spoken data transcribed for the current corpus uses the transcription system established by previous scholarship on North African dialects in order to provide comparability with existing work on the dialect; the transcription is closest to that of Heath (1989, 1997, 2002). This system retains conventions of work in Arabic dialects including the use of a sub-letter dot to indicate an emphatic, or pharyngealized, consonant and use of a caron to mark palatal-alveolars (Mercier 1959, Harrell 1966, Heath 1989, Ziamari 2009, Turner 2013, cf Sayahi 2014). Again following Heath, it is assumed that the vowel system in Moroccan Arabic contains three medial-length full vowels, the transliteration of which match their IPA sympols: /a/, /i/, and /u/. Ultra-short vowels are written as /ə/ and /ŭ/

following Aguadé (2010), although he notes that there is some variation in short vowels between dialects in addition to disagreement regarding their appropriate transcription. Figure 1.1, adapted from Turner (2013) gives the consonantal phonemes of Moroccan Arabic with the symbols used for them in this dissertation. The French found in the corpus is given in standard French orthography unless the word is an established borrowing in Moroccan Arabic (§5.4).

	(Bi-)labial	(Denti-)alveolar	Pharyngealized (denti-)alveolar	Palato-alveolar	Velar	Uvular	Pharyngeal	Glottal
Voiceless stops		t	ṭ		k	q		ʔ
Voiced Stops	b	d	ḍ		g			
Voiceless fricatives	f	s	ṣ	š		x	ħ	h
Voiced fricatives		z	ẓ	ž		ɣ	ʕ	
Nasals	m	n						
Laterals		l	ḷ					
Rhotics		r	ṛ					
Semivowels	w			y				

Figure 1.1 Moroccan Arabic consonantal phonemes

2. Language in Morocco

The four most common languages in Morocco are Moroccan Colloquial Arabic (MCA), Standard Arabic (SA), Tamazight and French. These four are used in different aspects of everyday life, with MCA and Tamazight being the primary languages of informal communication. Spanish and English are also used in certain regions and domains, but play a smaller role in the country. In this chapter, MCA and French will be detailed to a greater extent than the other languages as code-switching between these two language varieties is the focus of the dissertation.

With many languages used in the country, a logical question is why focus on the use of French and Arabic? And why distinguish between varieties of Arabic? The answer to these questions are linked as the answer to each relies in part on the concept of *diglossia*. First discussed by Ferguson (1959), diglossia refers to a situation in which two related language varieties are found in the same community, but are used in separate domains due to ideological reasons (see §2.2 for additional information). In Morocco, MCA is the 'low' variety of Arabic, used primarily in informal communication. MSA is the 'high' variety used in formal settings. As CS is typically found within informal settings, it is the low variety that is switched with French. Spanish is also found in CS with MCA (Sayahi 2007, Vicente and Ziamari 2008) and the increasing prominence with English has lead to a low level of CS with MCA evidenced in this corpus, although the extent of CS with English has not yet been documented.

The complexity of the linguistic situation cannot be understood simply by considering which languages are official. For this reason, a discussion of the history and roles of these languages is found in this chapter. A description and historical outline of each language is presented in §2.1. Attention then turns to the interplay between these

languages with consideration of diglossia in §2.2. Language policy and its effects on language use are described in §2.3. A linguistic portrait of the target community is found in §2.4, followed by conclusions regarding the relevance of code-switching to the contact setting in §2.5.

2.1 LANGUAGES PRESENT IN MOROCCO

Only two of the six languages spoken in Morocco are native languages of the country: MCA and Tamazight (often called Berber). SA, the third language from the Afro-Asiatic family, is the primary language of education and introduced in the first year of formal schooling. The two native languages are discussed first, beginning with MCA in 2.1.1 and followed by Tamazight in 2.1.2. Standard Arabic, the longest-standing official language of the country, is the next focus in 2.1.3. Proximity and trade between Europe, particularly Spain and France, and Morocco has long been a source of contact between cultures and languages of each country. This connection has been strengthened through colonial actions taken by Spain and France in Morocco and continues today through trade and media, including television, movies, music and internet communications. The European languages present in Morocco are considered in order of their relevance to the current study, beginning with French in 2.1.4, then English and Spanish in 2.1.5 and 2.1.6 respectively.

2.1.1 Moroccan Colloquial Arabic (MCA)

The most commonly spoken language in the country is MCA, used by 89.8% of the population over 5 years old (Haut Commissariat au Plan 2004). MCA has no official standing and is not standardized, nor is there a push for creation of a standard written or spoken form among speakers. Most of its speakers view it as a 'dialect' that is inferior to

Classical Arabic (CA), the Arabic of the Qu'ran, which they often view as the linguistic ancestor of MCA. This inferiority extends to comparisons of MCA with the dialects of other regions, as MCA is commonly considered incomprehensible by other Arabic speakers (Hachimi 2013). Like most other Arabic dialects, there is no literary tradition in MCA. However, the dialect has recently come to be used in many television series and is the language used for dubbing numerous foreign soap operas.[4] Music in MCA is also popular and the majority of music by Moroccan groups and singers is sung in the dialect (Caubet 2008), such as Ahmed Chawki who garnered international fame for his collaboration with American rapper Pitbull on the English/MCA bilingual song "Habibi I love you." The title of the song seems to indicate a strong influence of English with the only Arabic portion the Arabic term of endearment *habibi*, "my dear." However, there are only two phrases used in English by Chawki, seen in the chorus in (1.1)

1.1)	ħabibi, I love you	My darling, I love you
	I need you, ħabibi	I need you, my darling
	ɣanni-li w xali-ni mʕk	Sing to me and leave me with you
	ħabibi, I love you	My darling, I love you
	I need you, ħabibi	I need you, my darling
	duwweb-ni f nar ħbak	Melt me on the fire of your love

Although listeners who do not speak any dialect of Arabic are unlikely to understand the MCA portions of the song, speakers of other dialects are exposed to MCA through the lyrics of Chawki and many other artists. Despite the growing use of MCA in artistic endeavors, the dialect is not officially used in formal education where SA, one of the country's two official languages, continues to be the language of instruction. This is likely

[4] Egyptian Arabic is an exception, with full novels sometimes found in the dialect (Eid 2002). The Egyptian Arabic version of the popular site Wikipedia contains over 12,600 articles.

due to ideological factors, which have also prevented the standardization of written MCA (but see §4.3 for detailed discussion of writing non-standard languages and the emerging MCA writing system.) Yet studies have found that MCA is used commonly in spoken communication between students and teachers (Ziamari 2008) and may be used in a deliberate effort to increase comprehension of subject materials (Boutieri 2010).

Research into the origins of MCA reveals the high level of contact that has always accompanied the dialect, which has not simply experienced a gradual divergence from CA as is commonly believed. The first Arabic speakers in Morocco, laying the foundation for MCA, likely arrived in the second wave of Arab conquest of North Africa in the late seventh century (Naylor 2009, Magidow 2013). However, this group was not large enough to force the indigenous Tamazight population of Morocco to shift languages. Heavy waves of Arabo-Islamic migration in the 9th and 11th centuries assisted in the spread first of Islam and later Arabic, perhaps due in part to the upward mobility available to speakers of Arabic and societal prejudice against Tamazight speakers. In addition, the 11th century marks the beginning of the Christian "reconquista" of Andalusia that prompted emigration to North Africa by speakers of Arabic fleeing Spain, which ended with the expulsion of the Jewish and Muslim populations at the end of the 15th century. All of this demographic movement makes it difficult to ascertain the exact sources of the dialect spoken by Moroccans today, while its historical form and development cannot be reconstructed.

The use of a single name for MCA does not mean that it is a unified variety. Without any pressure to speak a standard variety, considerable variation can be found in the MCA of different regions (Caubet n.d.:1). Still, such variation is known and understood; it is a point of discussion among Moroccans and some variation is present in television programs. Most studies of MCA focus on the emerging national koinè, and for that reason this is the reference variety of the current study. The structure of MCA differs in important ways from

that of SA, the most complete descriptions of which can be found in Caubet (1993) and Harrell (1962), but certain aspects of the grammar remain contested (Turner 2013). Relevant aspects of the grammar are discussed, as necessary, throughout the dissertation.

2.1.2 Tamazight

Tamazight is an Afro-Asiatic language that is found across North Africa, with the highest concentration of modern speakers in Morocco. Tamazight is the oldest language of the region with archaeological evidence pointing to its presence as early as 800 BC (Kossman 2012) and its speakers are typically considered the indigenous population of the region (Naylor 2009). Estimates of current speakers in Morocco range from 25% (HCP 2004) to 45% (Stroomer 2008). Three distinct varieties of Tamazight are spoken in Morocco, each associated with a given region: Tashelhit in the south, Tamazight in the Central mountains, and Tarifit in the north.

Despite the fact that the language was present before Arabic, Tamazight only gained official status in 2011. Like many other indigenous languages, it has been marginalized and was long denied official recognition: it was absent from the 1956 constitution of the independent Moroccan government. Tamazight groups fought for recognition of their native language and it was slowly granted; in 1994 short news broadcasts began in the three primary Tamazight dialects, and the king decreed that the language would be used in formal education. The latter took nine years to be realized, as the language was introduced in 300 elementary schools around the country in 2003 (Bassiouney 2009). Tamazight was finally recognized as an official language of Morocco in the constitutional reforms of 2011. It is now available as an elective language in school and is found on an increasing number of official signs and information sources.

2.1.3 Standard Arabic (SA)

Standard Arabic has been an official language of Morocco since the country gained independence from France in 1956, and was the sole official language until Tamazight gained recognition in 2011. SA is the language of compulsory education in Morocco and is introduced to students from the first day of formal schooling. Official status does not automatically create a high level of skill in the language or insure its use in every day life: the most recent figures from the Moroccan government (2004) indicate that 56.8% of Moroccans over age five can read and write in SA, while a full 43% report not being able to read and write in any language. The World Bank's survey of 2000 households (2010) reports an increase in literacy with 62.5% of respondents able to read and write SA. The remaining 37.5% are illiterate. SA is the primary language of print media and, in name, the language of government and business. State-sponsored news is broadcast in SA in addition to Tamazight and French.

SA is an official language across Arab countries, including countries in which only a minority speaks the language. In this way it is an important part of pan-Arab identity and continues to be used in many domains, despite the fact that it is not the native language of any speakers: SA may be heard in the home through recorded Quran recitations, religious sermons, or news broadcasts, but it is not spoken to children from birth and is not used in informal settings. Native speakers of MCA often see SA as difficult, complicated and unsuitable for everyday use in speech (Bentahila 1983, Ennaji 1991)

2.1.4 French

Despite a lack of official status today, French has the greatest presence of any European language in Morocco. It is a compulsory subject in school from age 8 and is considered to be the 'elite language' (Ennaji 2002, Bassiouney 2009). Yet this emphasis on French does not guarantee widespread proficiency in the language, and reported French

literacy figures are declining. The 2004 census found that 39.4% of the population was able to read and write in French, but a 2012 study carried out by the World Bank indicates a decrease to 37.6% literacy in French. One difficulty with these surveys is that they focus on literacy and do not account for those who may be able to speak a language fluently or functionally, but who do not read and write in the language. French is common in many spoken situations, such as shopkeeper interactions with foreigners, even when users may not be literate in the language. This is in sharp contrast to SA, which is never used in informal spoken interactions despite its higher literacy rates.

The French government claimed Algeria as a colony in 1830, but showed little interest in Morocco until the Treaty of Fez in 1912, wherein the majority of Morocco was made a protectorate of France. Thus, the official, governmental use of French in Morocco dates to the time of the French protectorate. France's control of Morocco was short compared to other countries in the Maghreb and lasted 44 years, beginning in 1912 and ending in 1956. However, this control alone did not solidify the position of French as only 7% of Moroccans spoke French at the time of independence (Sirles 1999). The use of the French language during colonial rule is detailed in §2.3.1 below; here its current relevance is outlined.

Today, all Moroccans learn French through education, but the quality and efficacy of language instruction and curriculum varies greatly between education types, such as public and private schools, as well as between public schools because of teacher differences. French is a mandatory subject in school from the third year (8 years old) through the second-to last or final year of secondary education (17-18 years old), depending on the area of studies, with the majority of university subjects taught in French. Nevertheless, university teachers anecdotally report concern with their students' ability to use French for their studies and feel that they must frequently use Moroccan Arabic in the

classroom in order to ensure comprehension of course material (see Boutieri 2010 for a thorough investigation of language use in secondary education, see also Daoud 2011 for the Tunisian context). However studies by Chakrani (2010) and Chakrani & Huang (2012) on reported language use indicate low use of MCA or MCA–French code-switching in the classroom.

The contradiction between these findings is likely due to research methodology and target population: reported use is notoriously inaccurate when compared to observed use in situations of code-switching (Pfaff 1979, etc). Ziamari's (2008) findings illustrate this disconnect: in spite of negative opinions of CS, it is found in her classroom recordings even when the instructor is present. French is the expected language of the classroom, particularly in her participants' field of engineering, but MCA was regularly employed alongside it. When her informants later listened to some of the recorded conversations and classroom sessions, they were surprised that they code-switched. This indicates why reported use likely represents attitudes more than reality; Chakrani's youth may underreport usage because they believe these codes *should not* be used in the classroom, while teachers may report high levels because they feel that they use these codes more often than is proper. The teachers' concern for students' level of French, real or imagined, indicates an uncertain footing of French in Morocco.

In spite of concerns about French language proficiency among youth, French remains popular among students and professors. Ennaji (2002) found that university students continue to favor French; 73% of those he surveyed were in favor of French–Standard Arabic bilingualism, as were 78% of the professors surveyed. Their opinions are not surprising, given that most of the political and business leaders in France receive some or all of their university education in France (Sirles 1999). Its current value on the job market also cannot be ignored as employers prefer to hire individuals with advanced French

skills, increasing the 'linguistic capital' (Bourdieu 1982) of French. On a more practical level, France continues to be the largest trading partner with Morocco and is important in tourism as many tourists to Morocco are from France and those who are not may be more comfortable attempting to communicate in French than in Arabic. Tourism remains one of the largest sectors of the economy and the government is making efforts to expand it (see §2.3.3). All of these factors point to the privileged status that French continues to enjoy in Morocco. Code-switching between MCA and French is potentially an avenue through which speakers may access the prestige associated with French.

2.1.5 English

English is the most recent addition to the linguistic profile of Morocco. No exact figures exist on speakers of the language, but many studies point to its growing importance, particularly in culture and business (Ennaji 2002, Chakrani 2010, Bassiouney 2009). The use of English in education has grown and it is now offered as an option from age 16 in public schools, and it is taught as early as age 5 in private education. University content courses can also be found in English. One example is the field of linguistics, in which a textbook, written by Moroccan linguists for a Moroccan audience, whose main text is completely in English with examples in French, English, MCA, SA and Tamazight (Ennaji & Sadiqi 1992).

There is some evidence that interest in English may come at the expense of French. Recent studies have found that many young people are more willing to put effort into learning English than learning French (Chakrani 2010). This may be due to the fact that English is seen as an international language and does not have any connection with the history of colonialism. It may also be a reaction to the maintained importance of French by the elite of the country; young people who are against this trend may choose to index the

notion of modernity and access to technology through English as a different prestige language. Availability of curricular materials in French has been one argument for its privileged status in Morocco (Ennaji 2002), but if English continues to be used by young adults and in university classrooms, English could begin to compete with French.

2.1.6 Spanish

Spanish is spoken by "a considerable population" (Scipione and Sayahi 2005) in the north of the country. As Spain is Morocco's second largest trading partner, and in light of the mere eight miles of water that separate the two countries, it is no surprise that the Spanish language continues to be used in Morocco. While Spanish continues to be spoken in certain regions, it is taught, only as an option, in 42% of high schools in Morocco. Notable Spanish influence in Morocco began in the 15th century when the Spanish monarchy began the expulsion of Jews and Muslims. Many of the exiles went south and cities such as Fez continue to be celebrated for this early immigration. However, its current sphere of influence is limited primarily to areas controlled by the Spanish protectorate from 1912 to 1956 (Sayahi 2011b).

The presence of Spanish is found most often today in the regions that were once Spanish territory or protectorate, some of which continue to border Spain. The latter is the case in the north of Morocco where the two Spanish enclaves of Ceuta and Melilla exist on mainland Africa and are bordered by Morocco and the Mediterranean Sea. Spanish also remains common in the south where Morocco once bordered the Spanish Sahara (Sayahi 2011b). Now named the Western Sahara, the disputed territory is claimed by Morocco and independently as the Sahrawi Arabic Democratic Republic, whose government proclaims Spanish a co-official language with Arabic. While the importance of Spanish cannot be

denied in these areas, it does not play a major role in the center of the country where the current study was carried out.

2.2 DIGLOSSIA AND MULTILINGUALISM IN MOROCCO

The interconnected effects of language policy, social factors and economic concerns make it hard to understand the roles that each of these languages plays in Morocco. Morocco is clearly a multilingual country and, like any multilingual setting, the languages and varieties do not enjoy equal status. Societal preference for one language over another is often discussed in terms of prestige and domain of use: languages with overt prestige are used in more formal settings, such as government and education, while languages with covert prestige are used in informal or in-group communication. French and SA continue to carry overt prestige in Morocco, while MCA carries covert prestige (Bentahila 1983, Ennaji 2005, Chakrani 2010).

Prestige and domain of use are at the heart of Ferguson's (1959) classic description of diglossia, a term that is often invoked in discussions of Arabic as it is one of the languages he used to illustrate the term. He defined diglossia as:

> "a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation" (1959, 336)

This definition includes two varieties of one language: the 'high' variety is a highly codified, written variety, while the 'low' variety is spoken in ordinary conversation. Much of this definition can be clearly applied in the Moroccan context: as in other Arabic-speaking countries, SA or Classical Arabic present as a highly codified, grammatically

complex variety used for education that has a respected body of written literature and would typically be identified as the 'High' language of a diglossic situation. By contrast, it is doubtful that SA is used for 'most written and formal spoken purposes' in Morocco due to the presence of French. The complexity of language contact has lead some authors to expand the concept of diglossia to include unrelated language varieties or middle varieties that result from CS between the high and low diglossic varieties (Fishman 1967, Romaine 1989, Youssi 1995).

I qualify the language situation in Morocco as diglossic between levels of Arabic while simultaneously multilingual in order to take into account the other languages present, following Sayahi (2014). Walters (2003) suggests that the situation described by Ferguson may have been more idealistic than reflective of use in any Arabic-speaking country from the beginning and describes Tunisia as a post-diglossic society due to the many different varieties of language used. Sayahi (2007, 2014) uses a strict diglossic distinction between SA and TCA to help illustrate why CS in Tunisia occurs commonly between French and TCA, but not between French and SA in Tunisia. The recognition of both diglossia and multilingualism in modern Morocco provides a distinction between the special relationship found between SA and MCA while also acknowledging the roles of the other languages within Moroccan society.

2.3 LANGUAGE POLICY AND PLANNING IN MOROCCO

The impact of the language policies of France, Spain, and independent Morocco has helped to shape the current linguistic situation described above. French and Spanish colonial language policies are discussed first, as significant differences in the colonizers' approaches to educating the local population affected the status of each language on

independence. The policies of independent Morocco are then considered to understand how an independent government with Arabic-only as its ostensible goal has encouraged the use of French in particular.

2.3.1 French policy during the protectorate

In 1912, the French protectoral government established French as the sole official language of government, including government-sponsored school, and the media. French was the only language used in formal education and fluency was required to work in the government. Yet the French government did not create a bilingual Moroccan population, and this may not have been their goal. A four-part educational system was created that encouraged certain groups to learn French and become part of the administration, while casting others as manual laborers. Jewish students of both sexes were given schooling primarily in French, with Hebrew taught as an additional language. Non-Jewish Tamazight schools used French as the medium of education in an effort to encourage ties with France[5]. However, for Arab Moroccans, all presumed to be Muslim, education was divided by sex, status and region. Male students from the colonialists' chosen elite group were eligible to attend more rigorous schools that required tuition but supplied better French training. This 'elite' group of students was chosen from the sons of nobles and were expected to join the French administration after graduation. Beginning in 1930, students in these schools could earn the diploma necessary for entry into university studies. In the schools created for male Arab students from non-'elite' groups, education centered on providing the knowledge necessary to conduct manual labor in their surroundings: construction in cities, fishing on the coast, and agriculture in the interior. These schools used French as the primary language

[5] Berbers were extremely resistant to colonial rule. Berber schools educated far more Arabic students than Berbers, as Arab students enrolled in these schools to access the higher curricular standards. These schools were sometimes closed to due to the low proportion of the 'target' Berber population in attendance.

but the focus was on oral communication. As might be guessed from their limited scope, completion of these schools did not give access to higher education. Education for girls was limited to vocational and Qu'ranic training until 1938, when a primary-level certificate became available to female students (Knibiehler et al. 1992). These academic divisions divided the country by knowledge and opportunities, a division that has never disappeared.

Although a divided educational system was in place, the protectoral government did not make a great effort to spread language instruction to all Moroccans. At the end of the protectorate, it is estimated that only 10% of the non-Jewish children of Morocco attended a government school. A physical separation between Europeans and many Moroccans may have played a role in the low levels of attendance. The relatively small population of colonizers lived on the coast, while the majority of Moroccans lived inland. However, distrust of colonial schools and dislike for the curriculum available to most Moroccans also decreased attendance. Moroccans, dissatisfied with the limited colonial educational opportunities, created independent free schools to provide a higher level of education and continued to use existing traditional schooling, including those centered on Qu'ranic education. In Casablanca alone Moroccan-run schools enrolled 9,462 students while the French schools for Arab Moroccans counted 6,685 at independence (Benzakour et al. 2000). While the French government did not support or sanction these schools, their popularity indicates it was another way for students to get some kind of education when they could not enter French schools, or chose not to.

French instruction was included in the French government schools and the Moroccan free schools, but these efforts were not enough to create a broad population of French speakers. The first census of the independent government was conducted in 1960, four years after gaining independence, but just 7% of Moroccans reported the ability to speak French and only 6% reported the ability to read in it (Sirles 1999, Maroc Service des

Statistiques 1960). The lasting impact of French colonial policy was not in how many Moroccans spoke French, but which ones. This impact is discussed in 2.3.3 below.

2.3.2 Spanish language policy

Spain controlled two geographically distinct areas of Morocco during the early 20th century: Spanish Morocco in the far north from 1912 to 1956 and Spanish Sahara in the far south of the country from 1884 to 1975. In the north, Spain instituted a three-part educational system divided by ethno-religious groups. The branches were created to separately educate Spanish, Jewish, or Muslim students. The schools for each group were conducted in the language that the protectorate found most appropriate, leaving the Muslim school taught entirely in Arabic. As a contrast, the low population density in the Spanish Sahara lead to Moroccan students attending the same school as Spanish students, all of whom were taught in Spanish. Sayahi (2011b) attributes the relative maintenance of Spanish in each region, less in the north and more in the south, to the divergent language policies in place during the protectorate.

2.3.3 Language policy in independent Morocco

Arabization, the nationalist movement that focused on increasing the use of Arabic, began on independence in 1956. The primary goal of arabization is to convert education, government and public business to the Arabic language and distance Morocco from the influence of French government and language. The first minister of education declared that education should be in Arabic from the primary level (Sirles 1999). Increased access to education was highly valued by most citizens, and the numbers of students quickly increased. The independent government emphasized the importance of Moroccan instructors to teach in the new school system. However, local instructors had trained in French under the protectorate and continued to use French to teach their subjects. Foreign

teachers, mainly from Egypt, were hired in an effort to increase the use of Arabic. Although the language of education is shared between the two countries, the accents of the new instructors and their occasional use of Egyptian dialect presented comprehension difficulties for students. Outside of the language challenges, the Egyptian teachers brought more conservative political views, which clashed with those of Moroccans (Ennaji 2002). These early difficulties were followed by inconsistent policy over the next decade; as the minister of education and number of students in need of education changed, so did the level of arabization in schools, creating swings toward higher use of Arabic, and sometimes back toward French. It was not until 1989 that the educational system became arabized through secondary education. During the intervening thirty-three years Moroccans were required to learn French as both a foreign language and the language of a decreasing number of subjects, effectively creating a basic level of proficiency in both French and Arabic for an entire generation.

While arabization efforts have resulted in Arabic as the main language of education, the privileged status of French is maintained today through formal schooling. Standard Arabic is the official language of all content courses at all levels. Students also study Arabic language and literature for 6.5 hours per week early on, decreasing to 1-5 hours per week during the final two years, depending on the student's chosen specialization. French is the required second language beginning in the third year of schooling, approximately age 8, when it is studied for 8 hours per week. This is gradually reduced over time to 4-5 hours in the final two years of schooling. A second foreign language is only offered in the 10th to 12th years, seven years after French is first introduced. Despite its secondary status in compulsory education, the importance of French increases for the majority of subjects in

higher education. The few subjects taught in Arabic include Islamic studies, law[6], and certain social sciences. French remains the language of instruction for the more prestigious fields: medicine, engineering, hard sciences, law and most social sciences (Bassiouney 2009). The student who has not mastered French by the completion of secondary education is unlikely to be successful in university studies conducted in it.

Perhaps the most important factor in the maintenance of French is entrenchment of the language among the elite of the country. The elite who had adopted French during the protectorate retained their privileged status in independent Morocco. As gatekeepers to certain fields of employment, they made proficiency in French a prerequisite for entry into desirable fields of employment. Recent research on language attitudes in Morocco points to the continued importance of French in business (Ennaji 2002, 2005, Chakrani 2010). The elite of the country perpetuate this situation by sending their children to private schools that place emphasis on French thereby creating another generation that will become proficient in French. This allows companies to continue requiring French proficiency and, in turn, demonstrates the importance of the language in everyday life. As arabization efforts and enrollments increased, access to adequate French instruction decreased, likely caused in part by underprepared teaching staff. Subpar French courses in public education created an even greater distinction between those who could afford French education, and spoke French fluently, and the majority of the population who rely on public education. This distinction continues to make proficiency in French a valuable commodity.

Pressure to maintain French by the elite is a daily reality in Morocco. There are common stories of upper class families who go so far as to speak French, not Arabic, at home with their children in order to secure comfortable futures for them. Such families are

[6] Law is the only subject taught in both languages

not urban legends: while I was a student at an Arabic language school in a conservative Moroccan city, the school director joined us on a weekend excursion with his two young children. He and the staff from the school spoke to the children only in French. My Arabic teacher at the same school, who taught Arabic at a local high school during the academic year, stated that he supported the policy of arabization and the use of Arabic as the primary language of the country, but also wanted the best for his children. Part of his motivation to teach foreigners during the summer was to pay for private school to allow his children to speak better French so that they would have more career opportunities after graduation. An economic and social situation perpetuated by those who have benefitted from the current system seems to have secured for the French language a status greater than that given to it by the French government nearly sixty years ago.

Although the government has denied French official status and worked to decrease the amount of French used in education and many public professions, they have not quit using it. The government-sponsored news is broadcast in three languages: MSA, French and Tamazight. The multilingual broadcasts appear to be not only a sign of openness to French visitors, but also preferred by some Moroccans. For example, some government officials speak in French when interviewed for news problems, likely due to a higher comfort level in French than in Arabic. Such interviews are later dubbed into MSA for the Arabic version of the news; the same official does not simple repeat his or her statement. The government also acknowledges the practicality of using French for business. They wish to expand tourism, the second largest contributor to the GDP and second largest job creator in the country (Federation national du tourisme 2010). Emphasis is placed on foreign languages, including French and English, in order to reach the stated goal of doubling the size of the sector between 2010 and 2020.

The choice of French or Arabic in education, business and government has created tension among what has become known as the arabized elite and the francophone elite (Ennaji 2002). The francophone faction points out that the fields of technology and science are uniquely taught in French at the university level and thus, they argue, it is only natural that they should be taught in French at the secondary level in order to prepare students to succeed in these subjects. The francophone group also points to curricular factors: they contend that appropriate course materials do not exist in Arabic for many university-level fields and would need to be created before arabization could take place. Those that argue for the importance of arabization counter that the use of Arabic in education should extend through the end of university study, and stop the de facto disadvantage for students with weaker French. The existence of this ongoing argument indicates that while the languages may each be present in daily life, there is a certain level of awareness that their use is not neutral; a preference for one minimizes the importance of the other.

A government's policy undoubtedly impacts the way in which languages are used within the country, but the Moroccan situation informs us that a language does not need to be considered official in order to maintain a prominent role within a country. The moden presence of French in mandatory education and preference among the elite of the country ensure a prestige role for French regardless of the language's official status.

2.4 LINGUISTIC PORTRAIT OF THE TARGET COMMUNITY: MEKNES

Each city of Morocco is unique in its history and geography, which lead to modern cultural and linguistic differences. For this reason, it is important to understand the setting in which I completed my fieldwork and collected the data for the current corpus. Meknes is located in the interior of the country in a region of fertile lands between the coast and the

mountains. Today, it is the eighth largest city in Morocco (Haut Commissariat au Plan 2004) with an estimated population of 750,000. Although it is a former seat of the Moroccan government and its history spans more than a thousand years, close proximity to Fes, a UNESCO world heritage site, means that Meknes is not the primary tourist destination in the region. The city is by no means isolated from the effects of tourism and international companies; it remains on tourist circuits, hosts several foreign call centers, and is home to a number of foreign residents. French in particular remains prominent in Meknes as one of the five French private schools in Morocco run by the French government is found here. At the same time, French is less visibly present in Meknes than other large Moroccan cities due the lower number of tourists and international companies. In addition to the presence of European languages, many individuals of Tamazight heritage live in Meknes. Despite their origins, the majority of young adults in this study who mentioned Tamazight heritage reported a limited or passive knowledge of Tamazight due to growing up in a primarily Arabic-speaking city. Chakrani's (2010) results indicate that Tamazight is increasingly associated with the lower classes, which may encourage youth to use it less often.

2.5 RELEVANCE TO THIS STUDY

This chapter gives an overview of the six primary languages used in Morocco covering how each came to be used in the country and its current status. Moroccan Arabic and Tamazight are the two primary languages of everyday communication, but French is increasingly found in this domain as well. Standard Arabic is used in a restricted domain, rendering it less susceptible to contact effects with each of the varieties except MCA. English and Spanish, while significant, are spoken and understood by a smaller proportion

of the population due to the lower level of education in these languages. MCA, Tamazight and French are therefore the languages that display the greatest contact effects due to the high level of proficiency in these varieties across Morocco and the less restricted domains in which speakers use them.

While many contact phenomena may be found between these languages in Morocco and other settings where multiple languages are in intense contact, the focus of the current study is on code-switching between MCA and French. CS may be found between a variety of combinations of the languages listed, but MCA-French CS was selected due to its relatively high level of documentation. At the same time, the existing research on this language pair often draws on isolated examples, a methodology that may be used to identify possibilities in structural aspects of CS, but cannot indicate the prevalence of different structures or connect these structures to extra-linguistic factors. A careful quantitative analysis can be compared with past qualitative results while providing original insights that cannot be drawn from qualitative studies. Quantitative findings may shed light on larger theoretical questions in the field of CS due to the typological distance between MCA and French (see §3.3.)

The sociohistorical background provided here is also key to understanding how and why speakers employ their linguistic resources. French has remained important in Moroccan society due to the economic and social factors noted above, but the way in which it is used is reported to differ among speakers of different backgrounds. In this dissertation I examine some of the factors that are suggested to impact use of French (see §3.2) in an effort to determine whether these impressions are supported by a statistical analysis of naturalistic language use. The given language background can aid in the interpretation of significant differences in the use of French by different subgroups of speakers, or may

explain why such differences do not exist, should this be the case. may then suggest future avenues for research in speaker attitudes.

3. Linguistic code-switching

Aspects of code-switching have been studied for a century (Espinosa 1911). The topic remains all the more relevant today as we recognize a wide variety of linguistic situations in which CS is found and we utilize different empirical techniques that allow us to investigate various aspects of these situations in different language pairs (Gullberg, Indefrey and Muysken 2009). Like any aspect of language use, CS is an ever-changing phenomenon that evolves over time and manifests distinctions within separate speaker groups. As Bentahila & Davies (1995) show, significant differences in the use of CS occur within a single generation[7]. Yet most studies continue to focus on a single facet of CS, such as the relative amount of languages used or the syntactic categories involved in switches. With notable structural changes in CS occurring in the span of a single generation in the North African context, it is important to examine language use today and identify the social factors that play a role in a given community or speaker group in order to provide a comparison with future speaker groups. This information will allow us to understand how the implications of this practice, common in many communities throughout the world, may change over time. This chapter describes the variables under study and justifies their use to create a frame of reference for the study that follows. These topics are explored in a broad context as well as more specifically in Arabic-French contact in order to situate this study relative to other studies on Arabic-French contact, as well as in the broader conversation in the field of CS and language contact.

Before exploring past studies of CS, a working definition of the phenomenon must be established. Much debate surrounds the definition of CS, including whether it is necessary to distinguish this phenomenon from the practice of linguistic borrowing. Some

[7] The results of their study and its significance are discussed further in 3.2.3 below.

authors contend that they are distinct phenomena and illustrate the potential theoretical implications of making this distinction. Others contend that they are related phenomena on a single continuum and thus that attempts to distinguish between the two are misguided. This issue is addressed more fully below (§3.1) and the definition of each phenomenon as used in this study is specified. Once the terminology as used here is established, §3.2 addresses social factors that are often implicated as relevant in the use of CS in a variety of linguistic settings and language pairs. Section 3.3 provides a brief discussion of structural aspects of CS with an emphasis on the study of Arabic-French in particular. The latter two topics are brought together in §3.4, which illustrates how social factors have been assessed via structural observations.

3.1 CODE-SWITCHING AND LEXICAL BORROWING

The definition of CS varies from study to study, to the extent that a recent handbook on the topic gives a broad definition, "CS is the ability on the part of bilinguals to alternate effortlessly between their two languages" (Bullock and Toribio 2009:2). The authors then provide examples of how CS may vary in speaker groups and research focus,

> First, [CS's] linguistic manifestation may extend from the insertion of single words to the alternation of languages for larger segments of discourse. Second, it is produced by bilinguals of differing degrees of proficiency who reside in various types of language contact settings, and as a consequence their CS patterns may not be uniform. Finally, it may be deployed for a number of reasons: filling linguistic gaps, expressing ethnic identity, and achieving particular discursive aims, among others. (p.2)

Recognizing the wide variety of related phenomena, CS is defined here as the use of lexical items from two languages in a single turn or utterance. This includes both intra-sentential switching within a clause as well as inter-sentential switching between clauses. Intra-

sentential switching is the primary focus of the current study due to its frequency in the target language community and its relevance to grammatical theory.

With a working definition established, CS can and must be considered in relation to linguistic borrowing, (the term borrowing will be used from this point on). Both phenomena are characteristic of language contact, and borrowing can be loosely defined as a word from one language becoming part of the other language. In more formal terms, a lexical borrowing is a lexical item W_x originating in Language L_x that becomes lexical item W_y, a part of Language L_y. The concept of borrowing is most clear when we speak of what are sometimes called established loan words, or words that are clearly, doubtlessly part of the recipient language due to their accessibility to monolingual speakers.[8] In a highly codified language with a significant focus on lexicography, like English, established borrowings can be found in the dictionary, such as *coup d'état* or *honcho*. Other foreign words that enter the English lexically in more transitory fashion might more readily be considered as code-switches even if they are understood by many speakers, such as *merci*.

The relationship between the two contact phenomena of borrowing and code-switching is one area of language contact research in which consensus has not yet been reached. Poplack and Sankoff (1984) and Poplack, Sankoff & Miller (1988) maintain that an important distinction exists between borrowing and CS, based primarily upon morphosyntax. This view is defended anew in Poplack's more recent work (Poplack 2012). Many other researchers place CS and borrowing on the same continuum as they find no satisfactory way of distinguishing between the two (Treffers-Daller 1994, 2009 Myers-Scotton 1992, Muysken 2000).

[8] This is most straight-forward in monolingual communities; any level of community-wide proficiency in more than one language immediately complicates the notion of borrowing.

Researchers from both points of view tend to agree that borrowings are accessible by monolingual speakers of the recipient language, whereas CS can only be used by individuals who have achieved some level of knowledge in the source, or donor, language from which a word is borrowed. Those who argue against a strong distinction between the two phenomena point out that the identification of a given lexical item on the spectrum between the source language and the recipient language is often impossible except at the endpoints. As many have posited that code-switching plays a role in borrowing, it is understandable why many search for the distinguishing line between the two. However, as Heath (1989) notes, it may not be possible to find this line. For this reason Lahlou (1991) distinguishes between the two phenomena on a functional basis, identifying a lexical item as a borrowing if it is the only or primary word used for a concept within the community. As a native speaker and member of the Moroccan community, Lahlou uses his own judgment to make this distinction. This criterion for distinguishing between the two may aid in the categorization of French-origin words for which no MCA equivalent exists, but may mask any competition between native and French-origin lexical items, rendering his notion of borrowings relatively conservative and impossible to replicate precisely.

Researchers who are non-native speakers often must rely on dictionaries for gauging the degree of entrenchment of a potential borrowing, despite the many limitations in doing so. Sayahi (2014:127-128) notes the conservative nature of dialectal Arabic dictionaries; they are typically created for foreign learners and are therefore often limited to forms of Arabic origin. This seems to be true of Harrell's (1966) Moroccan Arabic dictionary, which was written as a resource for university students and other adults with a need to communicate in the local dialect. For this reason dialectal dictionaries are often limited in scope to the types of interactions that foreigners might have with native speakers and therefore omit many topics, including those that might be associated with a European

language. Such dictionaries are published infrequently, with only one English-Moroccan Arabic dictionary (Harrell 1966). Although multiple editions of this dictionary exist, the content has never been expanded from the original edition. While this is one of the most comprehensive Moroccan Arabic dictionaries available, it was considered only a starting point by its first compilers (ix). A recent Spanish-Moroccan Arabic dictionary exists (Aguadé & Benyahia 2013) and certain common terms that clearly have a non-Arabic origin have been added, such as *bortabl*, 'cell phone,' but the majority of entries are strikingly similar to those in Harrell. The only available French-Moroccan Arabic dictionary dates from 1959 (Mercier) and is therefore silent on more contemporary technological terms. The conservative nature of these dictionary sources guarantees that any list of established borrowings assembled from them is similarly conservative and likely outdated.

One study of borrowings has added to the bilingual dictionaries to incorporate a more recent range of MCA lexicon, including potential words of French origin. Heath (1989) examined vocabulary in multiple regions of Morocco to identify words of French and Spanish origin and completed the most detailed investigation of borrowing in MCA. He defines a borrowing as a lexical item that satisfies phonological, canonical-shape and morphological rules of the borrowing language (23). At the same time, he acknowledges that this definition is idealistic and in practice is complicated by 5 major difficulties summarized here as:

1. Phonological differences between MCA and French are not as clear as one might expect.
2. Morphology is not always necessary on certain words, such as adjectives and adverbs, removing a potential indicator of language membership.

3. Affixation from MCA may occur spontaneously with an other-language item due to high frequency of the process.[9]
4. Stems and paradigms from an outside language may become minor inflectional paradigms, such as the English adoption of Greek *data/datum*.
5. High proficiency bilinguals may use French forms where an Arabic borrowing is well-established.

With these difficulties in mind, Heath notes that he classifies certain examples as code-switches, such as seen in 3.1, even though they are uncommon or used only by highly educated bilinguals,

3.1) ma-ta-y-t-***u:tiliza***-w-š
NEG-PRES-reflexive-3rd-utilize-PL-NEG
'They are not utilized' (Heath 1989:24)

Conversely, Heath notes that certain common words, including /*lagaṛ/,* 'la gare' or 'the train station' are included in the list of borrowings despite the fact that their phonetic form and definite article are French.

Heath is not the only one to highlight the complexity of identifying borrowings in MCA. Sayahi (2014) reiterates the common difficulty of Heath's fifth point regarding the form of established borrowings in language contact situations with continuing access to the source language, such as in North Africa. As French is required in education, the phonetic form of a French borrowing, in particular, may vary widely dependent on speaker preferences. This is anecdotally seen in Morocco, although it has not been specifically studied. For example, one speaker may tell you the word for 'fruit' is pronounced /disir/ while another insists that it is /desɛʀ/, phonetically faithful to the French source. A similar phenomenon is found on the other end of the spectrum; Ennaji (2005) suggests that

[9] In the case of verbs, Poplack would likely qualify instances of this phenomenon as 'nonce-borrowings'

borrowing is socially acceptable in Morocco, which may allow speakers with a lesser knowledge of French to be comfortable incorporating the few words that they do know.[10]

In an effort to account for these difficulties, Heath worked with informants from three cities to identify a large set of borrowings from Classical Arabic, French and Spanish. This investigation provides key background to the current study because he consulted multiple informants, including from the city of Meknes, and noted details such as regions of use and semantic specification or drift that had occurred in the borrowing process. Nonetheless, this survey was carried out before many of today's technological devices (i.e., laptops, tablets, mobile phones) came into widespread use in Morocco and for this reason is incomplete in these fields. In the present study, the use of a single French word, or a multiword expression designating a single-referent such as *salle de jeux* 'game room,' is considered a French-origin lexical item if it is absent from Heath (1989) and the dictionary sources noted above. The term 'French-origin' is used to highlight the uncertain language classification of the lexical items. The use of French-origin nouns by MCA speakers is taken up further in Chapter 7.

The impact of the lack of identified borrowings in technology is important in the understanding of Arabic-French CS because borrowings are common in fields where terms must be created. Certain created terms may have a logical meaning based on other words in a given language: in English a computer is a thing that computes. Some languages have borrowed the English term as computers entered general use in areas in which the language was spoken, whereas others, including French and SA, have created their own term calqued on the original: *ordinateur* in French and حسوب /ħasuːb/ in SA have the same basic meaning.

[10] As an anecdote, my first host mother remembered none of the French from her schooling, but made her high school-aged daughter laugh by saying *jamais de la vie* ('never in my life'), common in soap operas and music, when she was surprised by something that she considered to be in bad taste. Although she and her daughter appeared to consider this phrase clearly French, it was accessible to her due to the high degree of societal bilingualism.

These neologisms are not always accepted by native speakers of a language. In Arabic, حسوب has not become the popular term for computer in all regions, with many dialects adopting the English or French term, depending on the major European language influence. As technology was generally introduced in French, habit may be the cause of the continued use of this language in Morocco in online settings. This may be one reason that many Moroccans report using the Facebook interface in French, despite the fact that it is now also available in Arabic. Repeated use of the French buttons may have allowed French words such as *j'aime, partager,* and *invitation* to enter the Moroccan lexicon in the context of the website, as evidenced in the following comic:

(Man at door: Let me come in so I can click ***(I) Like*** and leave. Others: Liar… You want to click ***share***) Source: http://montada.echoroukonline.com/showthread.php?t=228074

In the current study the term *established borrowing* is used here to designate words that have been previously attested in Moroccan Arabic dictionaries, including Heath

(1989). The remainder of the items found in the data will be considered instances of CS for the primary analysis in Chapter 6. This is a conservative measure, but is chosen as it can be verified and reproduced by other researchers of any language background. In a context such as Morocco, where functionally monolingual speakers may have varied knowledge of a contact language, it is important to remain aware that definitively classifying a lexical item as belonging to Moroccan Arabic or French may be impossible if we use criteria that are applicable only to a given token. However, the possibility that French nouns could be borrowings is further explored in Chapter 7.

3.2 EXTRA-LINGUISTIC FACTORS

Code-switching has long been found to be associated with social factors. Blom and Gumperz (1972) point out that setting, situation and identity all play a role in the use of CS between the standard language and dialect in Norway. Poplack (1980) illustrated for the first time that the grammar of CS may be impacted by competence in the relevant languages, and also found effects for sex and age of acquisition. Bentahila (1983) found evidence that topic and interlocutor are important in predicting whether Arabic-French bilinguals participate in CS. A wide variety of factors have been found relevant to the rate at which CS is used or to the structures that are most implicated in CS; some of the same factors are found by separate studies to affect both. These factors are also found to affect variation in monolingual language use; it may be that other factors that are often found relevant in monolingual contexts will also impact a bilingual's use of CS.

The impact of extra-linguistic factors on the use and structure of CS is so great in some cases that Gardner-Chloros (2009) suggests that they must be better understood before the syntactic patterns often observed in accounts of CS can be considered within a

given syntactic theoretical framework. In order to identify the factors that are most important to explore in a study of CS and to situate this study relative to past research, the following sections survey these variables and their impact on CS. Social class and the many implications it has in Morocco are discussed in §3.2.1. The variable of Language Attitudes is examined next in §3.2.2 as they are one of the first considerations in many studies on Arabic-French CS. The effect of Sex is overviewed in §3.2.3, followed by French Proficiency in §3.2.4. Finally, the speech settings in which CS is most likely to occur in the target community are explored in §3.2.5.

3.2.1 Social Class and language use

Social Class can be defined and measured in various ways, but always with the same goal: identifying the socio-economic status of the individuals it describes. While differences are certain to exist within such groups, this division has proven useful in many previous sociolinguistic studies as it allows for the identification of overall trends in the data. What is variously termed social class, social status, or socioeconomic status is often found to play a major role in language use. Labov‘s (1966) ground-breaking study of /r/ in New York brought to wide attention the impact of social class as a factor in sociolinguistic variation, and factors related to social class have since consistently been found relevant in accounting for stratified linguistic variation. Although it is often found to be a significant variable in sociolinguistic studies, very few studies on Arabic-French CS consider this factor. This may be due to the difficulty of assessing social class as the typical western predictors of education level, occupation, and salary are not reliable indicators of class in many Arab countries. Habib (2010) found it necessary to have a deep knowledge of the target community to divide speakers by social class. She found that income and residential area were the best indicators of class in her data, with occupation and education level less

informative in identifying class distinctions. Father‘s occupation was used by Chakrani (2010) to divide his speakers into classes; however, this was not as beneficial in the current data.

As noted in Chapter 2, a higher social class is historically associated with an increased use of French in Morocco, as well as an increased use of CS. Research often includes only speakers from a middle or upper class background in an effort to include more examples of CS. However, determining a given speakers' social class can be difficult when that person is not well known to the researcher. Chakrani (2010) notes that the traditional indicators of social class of America and Western Europe, such as property ownership or elite education in private schools, do not signify social standing in the same way in the non-Western context. Additionally, societal taboos often surround any inquiry regarding salary, and young adults, the targeted population here, are unlikely to know their parents' salaries. Job titles, when known by young adults, are frequently ambiguous. For example, an /ustad/ may be a university professor, an elementary school teacher, a language instructor at a center for foreigners, or a part-time tutor. It is impossible to estimate a social class without being able to identify the type of setting in which a person is employed.

Despite these difficulties, Chakrani utilized a participant's father's profession to identify social class and found that he had representatives from a range of social classes in his study on language use and attitudes in Morocco. He used this information to analyze the differences in reported language use among participants of each class in addition to their educational background (private/French-based and Arabic-based). He found that upper class participants used more French in classroom settings compared to their middle and lower class peers who preferred more MA-Fr CS. In an interesting connection, his matched guise results reveal that use of French is an indicator of the descriptor "*rich*" by speakers. In reported use, Chakrani found that participants from the middle and upper

classes, whether educated primarily in French or Arabic, used more CS with all family members. However, upper class participants also reported using more French with their brothers, and French-taught middle and upper class speakers reported a higher use of French with their fathers and sisters as well.

3.2.2 Language attitudes

Language attitudes in North Africa have long been studied, with the earliest research completed in the 1970s. Those early efforts, carried out in both Morocco and Tunisia, were concerned with how the language of the recent colonizers was used and perceived among the now-independent nations. As Micaud (1974) notes, the countries of the Maghreb were ostensibly aiming for arabization, but had instituted a bilingual school system that would maintain the importance of French despite the fact that French was not made an official language. Early studies found that higher knowledge and use of French was typically found in the upper classes who could afford private education and that the prestige of French continued to provide access to the most desirable jobs (Stevens 1974, Abassi 1977, Bentahila 1983). It is instructive to learn from recent studies that this situation has changed little in nearly 40 years as, despite broad changes in the educational system and increased efforts for arabisation, surveys from different Maghreb countries point to the similar preferences for French (Lawson & Sachdev 2000, Chakrani 2010, Sayahi 2011a, 2014).

Such language attitudes have been investigated directly via surveys as well as through Matched Guise Tasks. The earliest studies in the region used overt questionnaires to assess language attitudes in relation to reported use. The level of stigmatization of CS between Arabic and French during different periods in Morocco is unclear; Abbassi (1977) found that a large majority (78%) of his participants reported using CS, with 21% calling

it their favorite code for communication. This is both echoed and contradicted by Ennaji's (2005) findings. Most of his participants report using CS (70%), but at the same time 79% believe that the practice is "harmful to both languages" (p. 164).

Researchers have also evaluated speakers' preference for the language for a variety of topics and for media. In Bentahila's (1983) work, attitudes toward each language were gauged via the topics in which participants found them to be appropriate. Thus, he finds that French is highly associated with science and technical topics, while Arabic is used more often for religious topics and everyday interactions (p. 60-63). Lawson and Sachdev (2000) also found French to be associated with science and education, as well as social affairs. Their university-aged participants report using CS more than any other code, indicating a low level of stigma associated with mixing languages. As university studies are carried out in French in Tunisia, like for the majority of subjects in Morocco, university students may be more likely to incorporate French lexical items in their speech.

Reported use allows for a broad range of data to be collected quickly, but such results may reflect language attitudes and ideology more than actual use as speakers are often unaware of their linguistic habits, particularly when it comes to CS and may deny that they participate in it even upon hearing recordings of themselves (Blom & Gumperz 1972, Pfaff 1979, Lahlou 1993, Ziamari 2008). For this reason, some researchers have also employed Matched Guise tasks to probe Arabic-French CS behavior; two of these studies were based in Morocco (Bentahila 1983, Chakrani 2010), and one in Tunisia (Lawson & Sachdev 1997, 2000).

Lawson-Sako and Sachdev's (1997) matched guise results indicate that CS is rated the least positively of the five language varieties present in Tunisia on both status and solidarity dimensions by both male and female speakers. Other contrasts were found by sex; ratings for the male guises were highest for Tunisian Arabic, while ratings of the

female guises were similar among the other four varieties with the MSA guise rated as marginally higher than the French, English and TA guises. The authors also noted a lack of evidence of a diglossic relationship between the local dialect and MSA as the ratings for the two varieties of Arabic were very similar, suggesting that the speakers may see the varieties as simply different registers. In relation to topics, French and English were most associated with modernity and science, as expected, and both varieties of Arabic were associated with religion and tradition. The findings on topic then echo what was found previously through reported use.

Chakrani (2010, 2011, Chakrani and Huang 2012) reports similar results, with French related to education, modernity and being open-minded. The French guise in his MGT was also highly rated as desirable for a boss as well as a co-worker, revealing a continued connection between French and business, while the Moroccan Arabic guise was rated similar to French for desirability as a co-worker, indicating that the national dialect is valued to a certain extent in business. Standard Arabic was also rated highly for education, indicating that knowledge of a standardized, written language may be the key factor in sounding educated. The only characteristic for which MCA was rated significantly higher than SA or French was 'sociable.' The lack of distinction for other solidarity traits may indicate that there is not a strong diglossic distinction between Standard and Moroccan Arabic, and that French is competing with both in domains that were once reserved for Arabic. Chakrani suggests that Moroccans may accept the use of French in these ways because it is a part of local culture, and no longer seen by youth as the language of the former colonizers.

Viewed together, these studies suggest continued association of French with education, science and technology. The role of French in informal domains, such as within the home, may be increasing but research consistently points to higher use of CS with

friends than with family members. As time passes since the colonial period, French may be viewed less as the language of the colonizers, and more as an opportunity to find better employment and connect with other people. English has also entered the linguistic market of Morocco and may soon compete with French as the language of social and economic advancement (Bourdieu 1982, Ennaji 2005, Chakrani 2010). However, the fact that French has been recently found to be highly rated for solidarity traits as well as status traits may be a sign that French has become a positive facet of Moroccan identity.

3.2.3 Sex and language use

Sex is often related to the variable use of the standard form of one's native language, with women typically employing more normative language than men (Wodak and Benke 1997). At the same time, there is a common finding in sociolinguistic studies across countries that women also lead change. This apparent contradiction is well-documented (Labov 1966, 2001, Eckert 1989) but has not been consistently found in the Arab world. This is not to say that studies have not examined sex or the use of different varieties in Arabic-speaking countries, but the way in which these studies have been conducted does not always produce clear results. Ibrahim (1986) blames this confusion on the terms to refer to registers, or styles, of language, and contests that the terms 'standard' and 'prestige' should not be conflated, particularly in Arabic-speaking regions. He notes that females are often expected to use more Standard Arabic than their male counterparts, but that they often do not despite this expectation. This could be interpreted as going against the trend of female preference for standard language (J. Milroy 1981, Labov 1991), if one depicts MSA as the standard form based on Ferguson's (1959) definition of Diglossia. However, Ibrahim argues that Standard Arabic is not the only source of prestige; according to his views,

female speakers achieve prestige through the incorporation of European languages into their speech. Thus, he argues that referencing language styles as 'standard' or 'non-standard' fails to capture the complex relative prestige of varieties in the Arab world and that prestige alone, although often associated with 'standard' varieties, should be the focus of studies. Ennaji's (2005) view, as well, supports the substitution of the term 'prestige' for 'standard' in Arabic contexts. He states that females use more French and CS as a means to help "make their voices heard" due to the power brought by the greater overt prestige of French. These women then use French with their children in order to impart the same access to power and prestige and thereby increase their childrens' future opportunities.

Lawson and Sachdev (2000) explored language attitudes related to sex in Tunisia via a Matched Guise Task. They found that listeners gave higher status ratings to a male using CS than to a female doing so, despite the fact that females are commonly believed to use more French and CS. In a related experiment, researchers of each sex and two ethnic backgrounds, European and Arab, asked passersby a simple directional question in either French or Tunisian Arabic and noted the language of the given response. Among their findings is that CS was used most often in the responses to the Arab female researcher, whether she posed a question in French or TA, giving further evidence of a subtle association with CS and females.[11]

Other work in North Africa also questions whether women actually employ more French, either alone or through CS. Lahlou (1991) found that Moroccan men use more French, but attributes this use to the fact that men are more likely to have jobs and that through employment they interact with more French speakers, leading them to use French more often than women. However, Chakrani (2010) found no difference by sex in reported

[11] Responses regarding sex on its own are less clear: CS was used in 9% of the responses to the European female researcher when she used French, but in 21% of the responses when she used TA.

use of French or CS among his participants. Women are often found to over-report use of standard forms (Labov 2001) and if we, like Ibrahim, think more about prestige than standard, the interpretation of Chakrani's result becomes less clear. If females over-report their use of French, then they may in fact use less French than the male participants. At the same time, his participants reported using Arabic-French CS more often with their sisters than their brothers, again indicating a subtle association of the practice with women.

Ziamari's (2008) results also question the association with women. Female participants in her study stated that they prefer French over other languages, while males report a preference for Standard Arabic. Ziamari's perception of their use mirrors these attitudes, as the females use more French and are more faithful to a Parisian French pronunciation than males despite the fact that the males are capable of using the 'standard' accent when they so wish. However, when asked their attitude toward CS, the female informants were neutral to it or said it was evidence of a lack of competence in one language, while the males considered it a valid communication strategy and a tool for artistic creativity.

Instead of relying on attitudes and stereotypes, Sayahi (2011a) compared the amount of French used by Tunisian males and females in a one-on-one interview. He found that overall rates of CS were very similar by sex, but the trends seen in his identification of the syntactic patterns of CS differ by sex. The female participants did not use more French with him than males of the same education level, but they did use French for different parts of speech; males employed nearly twice as many intra-sentential switches and notably more complete clauses in French, while females employed more bare French nouns.

The differences in the use of CS by sex is not strictly about who may use the greater number of lexical items of French origin. It may be that the way in which females

incorporate CS in each dialect is more salient than the way male speakers use CS, either by the pronunciation or the constituent types employed by female speakers. If this is the case, it is potentially the *type* of French used in CS by women, and not the *amount*, that creates or feeds the perception that women use more CS. For these reasons, the differences between the sexes regarding the rate of French usage and the type of constituents most frequently switched must be analyzed separately in order to identify their how factors such as Sex contribute to their variation.

3.2.4 Language proficiency

The potential impact of proficiency in each language as a factor in CS is often clear at an intuitive level: it is impossible to use words or phrases from a language that one does not know. It is then unsurprising that proficiency in each language often affects a speaker's use of CS. Poplack (1980) suggested that certain types of code switches require a higher level of bilingual proficiency, with intra-sentential switches most indicative of balanced bilinguals in her study. Nortier (1989) echoes this sentiment and found that high proficiency was correlated with more intra-sentential and single-word switches, again assumed to be the most difficult type, among MCA/Dutch bilinguals in the Netherlands. At the other end of the spectrum, tags, extra-sentential elements and phrases are often assumed to be the simplest and most common. These are all examples of the structures that may be used by speakers of varying proficiency and may be completely independent of the amount of French that they employ in CS.

In Morocco, as in the rest of the Maghreb, level and type of education are closely related to French proficiency. As university courses are taught primarily in French, most individuals who complete university-level degrees have a stronger command of the language than those who complete only secondary studies. This generalization comes with

the caveat of area of study: degrees completed in Arabic would likely not bring a better level of French. The type of educational institution attended, described in Chapter 2, can also strongly impact an individual's French ability due to the differing emphasis placed on French.

Sayahi (2011a) tested the effect of level of education on CS using Poplack's distinction between inter- and intra-sentential CS. He found no significant difference in intra-sentential CS use among males with secondary and university-level education. Rate of CS instead emerged as a significant factor as speakers with a secondary education were found to use half the amount of CS that their university-educated counterparts used. While he notes that the CS of university-educated speakers is more 'diverse,' his discussion of structural aspects does not go deeper than Poplack's assumed measure of complexity. However, he provides an elaboration and quantification of the parts of speech that were switched by each group and these reveal important distinctions: the secondary-educated male group switches most often for bare nouns (55.2%), notably higher than the same category by university-educated participants, both male (35.6%) and female (39.9%). This is further evidence that a quantitative analysis of the grammatical structure of CS, and not simply rate, may distinguish differences between participant subgroups.

Bentahila and Davies (1995) do not take into account Poplack's inter- and intra-sentential switch type distinction and find striking differences between two generations of Moroccans by examining the constituents that each generation produced in French. They focus on differences in French proficiency resulting from type of education due to the scholastic changes that have occurred since Morocco gained independence. The older generation in their study was born between 1939 and 1951 and all speakers within the group completed university-level studies, which would have been conducted largely in French. The authors consider that the older group consists of balanced MCA/French

bilinguals. The younger generation, born in the late 1960s and early 1970s, was educated after Arabization policies were in place, and was found to be less proficient in French. The structure of CS varies greatly between these two groups. The older generation used a wide variety of CS types with full clauses more common than any other type of constituent (25.5%). By contrast, CS of the younger group contained a full 50% DP switches and full clauses or sentences accounted for just 5.6% of the data. Exposure to French, particularly in formal education, and its place in every day life, is vastly different for members of each group and is reflected in the structure of MCA-French CS that they employ. Although many studies on Arabic-French CS do not account for French language proficiency (Naït M'Barek & Sankoff 1988, Lahlou 1991, Lawson & Sachdev 2000, Ennaji 2005, Ziamari 2008), it is important to include this factor as it may reveal important distinctions in the use of CS.

3.2.5 Setting and interlocutor

It is common in many communities to speak one way at home with family and a different way when reporting events to a police officer or in a court trial. The individuals involved in a given interaction and its function also play a role in speakers' linguistic practices. Blom and Gumperz (1972) found that casual conversations among community members displayed more local features than official business interactions between the same individuals. Setting and interlocutor similarly impact the use of CS in Morocco. This has been studied primarily through reported use surveys (Bentahila 1983, Chakrani 2010, Ennaji 2005) although some research has also investigated observed use of different participants (Abassi 1977, Bentahila & Davies 1995, Aabi 1999, Lawson & Sachdev 2000, Chakrani 2010). While students and teachers believe that studies should be carried out only in French (Ennaji 2005, Chakrani 2010, Boutieri 2010) or in Standard Arabic (Chakrani

2010) individuals report using CS within the university setting and have been recorded doing so, sometimes to their surprise (Ziamari 2008).

Bentahila (1983) explored reported language use by location, interlocutor and topic in an effort to better understand how these factors impact language choice in Morocco. He found, to his surprise, that participants report using more French and CS with friends than with family members, and again more outside of the home than inside of it. The same is seen in Tunisia where respondents report commonly using CS with friends, followed by family members (Lawson and Sachdev 2000). The results of Chakrani (2010) and of Chakrani and Huang (2014) point to the same fact in Morocco today with speakers reporting higher uses of French with friends than with family members even though French is reported to be used increasingly in the home.

Chakrani's research on language attitudes and reported use also shows a shift in the domains in which French is found. Traditionally, MCA was the primary language of the home, while MSA and French were the languages of the domains of education and technology. His university-age respondents report using French within the home and in the street through CS; these results may reflect the attitudes identified through the Matched Guise Task. In that context, students do not rate French to be significantly different from MCA and MSA on three of the nine solidarity traits that he included in his study (p.121). In fact, French is rated significantly higher than MSA on four other solidarity traits suggesting that French now competes with both Arabic codes in domains where they are traditionally assumed to not overlap. However, MCA retains some distinctions for solidarity as it is rated significantly higher than French and MSA for the trait *sociable* and higher than French, but not MSA, for *honest* and *patriotic*.

3.3 STRUCTURAL ASPECTS OF CS

Structural aspects of CS are relevant to any complete theory of language because such a theory should be able to account for bilingual production as well as monolingual language. Early research identified clear structural patterns of CS and many authors posited structural constraints on the practice in line with a given syntactic theory (Timm 1975, Poplack 1980, Bentahila and Davies 1983, Woolford 1983, DiSciullo, Muysken and Singh 1986, Belazi, Rubin and Toribio 1994, Mahootian and Santorini 1995, among others). Many constraints were proposed in the pursuit of this goal, including the Equivalence constraint (Poplack 1980, 1981), the Government constraint (DiSciullo et al. 1986), the Closed Class Constraint (Joshi 1985), among others (see MacSwan 2009 for a summary of proposed constraints). Arabic-French CS has been analyzed by various researchers as part of this search because the typological distance between these unrelated languages were considered to provide a useful test of constraints that had often been formulated on the basis of typologically close, and/or related, languages. While many of the proposed constraints were descriptively adequate accounts of the data on which they were originally based, most structural constraints have since been demonstrated not to extend well to other language pairs and contact settings. The data from these studies reveals a high level of variation between different corpora of French-Arabic CS.

To address the incompatibility of theory and data that are found when CS is addressed at a macro level, Muysken (2000) suggests that different types of code-switching may be seen in different language pairs and contact settings. He identifies three primary types of CS: *insertional* CS, as in (3.2), *alternational* CS, as in (3.3) and congruent *lexicalization* as in (3.4).

3.2) xdəm-t f-waħəd *la société d'assurance*
Worked-1st in-one *the company of=insurance*
'I worked in an insurance company'
(Moroccan Arabic/French; Heath 1989:34)

3.3) maar ‘t hoeft niet *li- ʕanna ida šeft ana...*
but it need not *for when I-see I*
‘But it need not be, for when I see, I...‘
(Moroccan Arabic/Dutch; Nortier 1990:126)

3.4) voor gistere hebbe *jullie ‘t ingedeeld* wa
For yesterday have *you it subdivided* eh?
‘For yesterday you have subdivided it eh?’
(Ottersum dialect/Standard Dutch; Giesbers 1989:158)

Insertional CS is the “insertion of material (lexical items or entire constituents) from one language into a structure from the other language.” (3) that results in the structure A B A where A is the primary language of an utterance and B the secondary language. In (3.2) above, Moroccan Arabic is the dominant language of the discourse that provides the grammatical frame into which French constituents may be inserted. *Alternational* CS is “alternation between structures from languages” (3) such that neither language can be considered dominant. This is seen in MCA/Dutch switching as in (3.3) where neither language seems to provide more of the structure than the other. Finally, *congruent lexicalization* occurs when a shared language structure exists between two languages, and that structure may be filled with lexical items drawn from either language. Examples such as (3.4) show that the lexical items switched in *congruent lexicalization* need not be full constituents.

Muysken posits that each type is possible in its given context, and that the previous constraints on CS can all be classified within these types. He suggests that each constraint may be applicable in the situation from which the original data was collected, but may not extend well in other contexts. For instance, he remarks that insertional CS is likely to be found in a post-colonial setting in which speakers are much more proficient in one language than the other, while alternation may be found in more stable bilingual communities, where speakers have more balanced bilingual proficiency. Each of these situations could be said

to be true in Morocco (and many other locations), where upper class individuals display stable bilingualism while lower class individuals have strong proficiency only in their native dialect, but not in French. Congruent lexicalization is expected to occur between typologically similar languages whose bilingual speakers do not feel the need to separate, such as a dialect and a standard language. Muysken's typology is difficult to operationalize (Deuchar, Muysken and Wang 2007) but it provides a linguistic external rationale for why supposedly universal CS constraints cannot account for all the empirical data that researchers have culled from various language pairings.

Muysken's typology may address why two constraints, each proposed in part on Arabic-French data, make contradictory predictions. The Functional Head Constraint (FHC) proposed by Belazi et al. (1994) predicts that a switch may not occur between a functional head and its complement, based in part on Tunisian data. In response, Complement Adjunct Distinction (CAD), was put forth by Mahootian and Santorini, based on data from Bentahila and Davies (1983) from Moroccan Arabic, who remarked that switches can and do happen between functional heads and complements (see Post 2010 for a detailed comparison of the two constraints). Using a modified grammaticality judgment task, I found that speakers from each country, in fact, did show significantly different preferences for the syntax of CS within the determiner phrase, where the D acts as the functional head to the NP complement (Post 2010). The speaker judgments from these locations were not fully consistent with the predictions of the either the FHC or CAD but, importantly, the divergence in the CS preferences of the speakers of these two communities of the same language pairing is notable. Although Tunisia and Morocco are both post-colonial settings, speakers' intuition regarding the goodness of CS patterns differs significantly between the two populations despite the overall similarity between these two sites in terms of the linguistic systems involved and the cultural histories of these sites.

Language proficiency may also come into play as compulsory education in Tunisia incorporates French in scientific subjects during secondary education, while this shift away from Arabic occurs at the university level in Morocco.

The Null Theory of code-switching, described through the minimalist theory of syntax by MacSwan (1997, 2009, 2014 among others), and the Matrix Language Frame model put forth by Myers-Scotton (1993, 2002, Myers-Scotton and Jake 2000, 2009, among others) are dominant universalist accounts of CS structure today, but from very different perspectives. The Null Theory, adapting a generativist perspective, abandons constrains on CS other than the constraints that are inherent to the grammars of each of the languages of the pair. The MLF, a psycholinguistic model of CS, relies on the notion of nonequivalence in CS—that one language contributes the structure (the *matrix language*) into which other language elements (the *embedded language)* are inserted. It is important to note that the MLF is intended for *Classic CS*, "CS in which empirical evidence shows that abstract grammatical structure within a clause comes from only one of the participating languages" (2009:337) These models and those proposed before them have shaped the discussion and study of CS; without the debate that surrounds structural constraints on CS, the phenomenon would not be as well-documented as it is today.

The observed difference in Moroccan and Tunisian CS points to one difficult aspect of identifying universal constraints on the structure of CS. Variation in CS structure can be found between different language pairs, as in Poplack and Meechan's (1995) analysis of Wolof-French and Fongbe-French CS but it can also be found between a similar language pairs, like with TA-Fr vs. MA-Fr. There is also evidence that variation in CS exists within a single dialect: Bentahila & Davies 1995 found structural differences between the two generations that they studied, while Sayahi's results (2011a) suggest imporant differences in CS structure based on proficiency or education level. These findings are not new or

limited to Arabic-French contact; Poplack considered structural variation in CS within a single community in 1980, even though she delineated her analysis to a categorical difference between inter- and intra-sentential. The presence of variation within and between communities is what leads Gardner-Chloros (2009) to doubt the explanatory power of constraints, preferring instead to focus on the social factors that impact the use of CS. It is unclear why researchers have not come back to structural differences in CS as one type of evidence of the way social differences affect bilingual language behavior but the current study, with a dual focus on social and structural properties of Arabic-French CS, seeks to help fill that gap.

A wealth of studies on Moroccan Arabic-French CS provides evidence of the changing use and structure of CS in Morocco, which may reflect the retrenchment of French usage vis-à-vis the Arabic dialects because the youngest generations have been educated and begun to work within an increasingly Arabized system (Bentahila and Davies 1995, Ennaji 2005, Chakrani 2010, Ziamari 2008). At the same time, it is difficult to confidently compare and evaluate the findings of these study due to the varying methodologies and level of detail provided by researchers.

As in other language pairs, nouns are often found to be the most commonly switched constituent in CS generally (Van Hout and Muysken 1994, Muysken 2000, Myers-Scotton 2002, among others) and this finding holds in Arabic-French data. Abbassi (1977) found that noun phrases are the most commonly embedded constituent in CS, and that CS typically occurs from Arabic, as the base language, to French, but he does not quantify the use of French in his analysis. Other researchers have provided quantification of the rate of nominal structures (NS) used in CS; these are summarized in Table 3.1 below. Nominal structures, defined as bare nouns, noun phrases, and determiner phrases, were

chosen for comparison because their presence in many data sets provides sufficient data for comparison between studies.

Table 3.1 French nominal structures (NS) as a percentage of all French constituents in four Arabic-French CS corpora

Study	**Participants**	**% NS**
Naït M'Barek & Sankoff (1988)	Moroccans residing in Montreal	35.1%
Bentahila & Davies (1995)	Older generation	22.2%
Bentahila & Davies (1995)	Younger generation	62.8%
Redouane (2005)	Moroccans residing in Quebec	28.1%
Sayahi (2011a)	University-educated Tunisian males	59.0%
Sayahi (2011a)	University-educated Tunisian females	63.8%
Sayahi (2011a)	Secondary-educated Tunisian males	71.3%

The data in Table 3.1 points first to differences between broad speaker groups, as each study was carried out in a different country. Differences within subgroups are also visible in the case of Bentahila and Davies (1995) and Sayahi (2011a). The studies by Naït M'Barek and Sankoff (1988) and Redouane (2005) were carried out among bilingual Moroccans residing in Quebec, Canada. These speakers, all native speakers of Moroccan Arabic, live in a setting in which Arabic is a minority language, unlike in Tunisia and Morocco, the settings of the other two studies. Individuals in Quebec likely have high French proficiency, as do the older generation in Bentahila and Davies' study, educated in a fully Francophone system shortly after the country achieved independence. The younger generation in Bentahila and Davies (1995) and those in Sayahi's study were educated in primarily Arabized systems, with French present as a foreign language until high school. Within Sayahi's group, proficiency distinctions may also be seen; university graduates tend to have a higher level of French as it is the primary language of post-secondary studies. While the varied linguistic settings of the three studies likely contributes to the way in

which speakers participate in CS, the broad trend in use of nominal structures may be related to proficiency, further discussed in Chapter 6. Another possibility is that a lower rate of nominal structures is used in situations in which French maintains a more dominant presence, as it did for past generations in Morocco and continues to do as the dominant language of Quebec.

Other trends in the structure of CS have been noted in many studies that inform the current research. Bentahila and Davies (1983) report that CS occurs at every constituent level and that switching from Arabic to French is most common, but that switches also occur from French to Arabic. Nortier's (1995) comparison of Moroccan Arabic in contact with French and with Dutch reveals a distinction between the two: definite articles are often deleted in MA/Dutch CS, while they seem to be obligatory in MA/French CS. She concludes that the difference lies in the clitic character of articles in both languages. However, null articles are 'very frequent' in Ziamari's (2008:97) corpus of conversational MA/French CS and are but one example of the evolution of the phenomenon. In fact, Nortier's analysis addresses only the contexts in which two determiners are required in Moroccan Arabic (see §7.3); it is not clear whether articles on French nouns may have been deleted in other contexts. The variation between groups with regard to CS points to a potential role of structure in understanding how speakers switch between Moroccan Arabic and French. One limitation in using of these comparisons for a potential meta-analysis is that few researchers have used quantitative methods, providing instead only isolated examples of CS structure as present in or absent from a given corpus. These reports make it impossible to identify the prevalence of each structure in the data sets from which they are drawn, if there is any potential variation with a similar structure, or if that variation might be associated with a subgroup of participants.

3.4 FOCUS OF THE CURRENT STUDY

Studies that have examined social factors among speakers who code-switch typically focus on only one of two possible metrics: calculating either the rate of other language insertions or the relative ranking of switched morpho-syntactic constituents. Rate, often expressed as the percentage of words in each language, is a rough indication of how much CS a speaker uses, but cannot capture the type of constituents that are employed. If a speaker uses 15% French words all in complete sentences, this is certainly different than a speaker who also uses 15% but all as single words through her speech. Relative usage of each constituent type by speaker may shed light on differences between groups. For instance, Poplack (1980) examined the binary distinction of inter- or intra-sentential switches across groups of Spanish–English bilinguals, who differed in proficiency. Sayahi's results (2011a) indicate that a link might exist between rate and structure such that those who used less CS also appear to employ a greater percentage of nouns.

The current study incorporates both rate and structure of CS in order to achieve a more complete picture of speakers' production. In order to reliably identify the social factors that impact CS we must look at the use of CS within a large sample of the population or within a group in which some factors that may contribute to variation are controlled. Comparisons between the speech of individuals residing in countries where the same languages are encountered, or where similar dialects are in contact with the same language, may reveal distinctions between speaker groups that can be attributed to social factors. Similarities in CS practices may indicate the aspects of CS that result from a given languge pairing. The work of Sayahi and Bentahila and Davies shows us that social variables do impact CS use, a finding that calls for further examination of CS practices and structures within a variety of communities. Similarities in CS use by speakers of a single age group or community will indicate the characteristic use of CS in their community. This would be

especially true of communities in which CS is the unmarked communication mode. Conversely, systematic differences in CS patterns will indicate the subtle ways in which speakers enact and maintain group membership and identity.

The current study uses both rate and structure of CS to investigate three factors that have been shown to be relevant in the use of CS between Arabic and French in Morocco as well as other language pairs. Sex is analyzed in a categorical way (M/F) with no attempt to discern the subtleties of gender norms. It is expected that male and female speakers will use a similar amount of French in CS, as found by Sayahi (2011a) and Chakrani's reported use (2010), but that variation between the sexes will be manifested in how French is integrated into Arabic. Language attitudes are also considered as an independent variable that potentially contributes to the varition in CS behavior, with positive attitude toward French and CS expected to correlate with higher use of CS. The third factor, Proficiency (in French), is expected to show a relationship between lower proficiency and less use of French lexicon overall, with a proportionally greater use of nouns in comparison to other French constituents. These hypotheses and the methodology used to explore them are taken up further in Chapter 5.

4. Computer Mediated Discourse

The field of Computer Mediated Communication (CMC) plays a role in the current study due to the inclusion of Facebook IM as the means of collecting the written CS data. CMC is defined as "predominantly text-based human-human interaction mediated by networked computers or mobile telephony" (Herring 2007; 1). Computer Mediated Discourse (CMD) is a subfield within CMC studies that focuses on the language used. As defined by Herring, CMD "is the communication produced when human beings interact with one another by transmitting messages via networked computers" (Herring 2001; 612). Both terms are useful in understanding how participants use MCA-French CS in IM as IM is a specific mode of CMC, and the language produced through it can be analyzed using methods from CMD studies.

Although studies on CS in CMC are becoming more common (see §4.4), few studies include CMD data from Morocco. The studies that do are from the early days of the internet in Morocco (Berjaoui 2001, Atifi 2003) or include a relatively small amount of data from the country with no analysis of CS (Zoabi 2012, Elhija 2013). Studies on written Arabic-French CS are less common, with just one focusing on any Arabic dialect in contact with French (see Bentahila & Davies 2002 for discussion of MCA-French CS in song lyrics). CS is traditionally a spoken phenomenon owing to its association with informal communication across communities in which speakers use CS, while CMD remains primarily written today. [12] Like other Arabic dialects, there is no official written

[12] The role of audio and video in CMC is doubtlessly increasing, but remains secondary to written material.

form of Moroccan Arabic. With these facts in mind, how does an 'unwritten' language variety such as Moroccan Arabic find its way into a written medium?

Many studies focused on spoken CS with written CS data limited to planned discourse, such as literature (Valdès-Fallis 1977), music (Bentahila and Davies 2002), historic texts (Schendl 2012), traditional communication forms such as personal letters (Nurmi and Pahta 2012) or private diaries (Montes-Alcalá 2000). However, the rise in popularity of the Internet has facilitated informal written communication as well as the observation of phenomena found in this mode and sparked interest in the relationship between informal CMD and spoken conversation (Hinrichs 2006, Herring 2010). The informal nature of forums, chat rooms and two-person chats gives users the freedom to express themselves in whatever language seems most appropriate, often with less concern for 'standard' language use.

The presence of Moroccan Arabic-French code-switching in CMD is, of course, a relatively recent phenomenon. I unite data from a variety of sources in order to situate written CMC as a practice within a larger social and technological setting in Morocco. This chapter includes findings from studies of CS in written CMC in other communities where 'unwritten' language varieties can now be found in this mode. Because of the recency of Internet availability, there is relatively little work on written CS and even less so in CMD (Sebba 2012). Two studies on written CMD by Moroccans are discussed as background to this research: Berjaoui (2001) and Atifi (2003). Internet access in Morocco was extremely limited at the time of these early studies and a comparison of their findings with the patterns of language choice and use in CMD today may indicate a changing population of Internet

users in addition to changing literacy practices as speakers choose to express themselves through writing in their native vernaculars.

Multiple frameworks have been used to examine CMC and CS within it. When focusing on CS used in CMC, it is possible to highlight differences between informal spoken and written modes on a pragmatic level, centering on the functions of CS in each mode. Alternatively researchers may examine the types of CMC interactions that contain CS. Another informative approach is the application of New Literacy Studies (Gee 1996, Barton and Hamilton 2000) as employed by Lee (2007) in an analysis of bilingual text messages. This framework considers the literacy practices and events that are found in CMC as a new environment for language use. Findings from each of these approaches indicate that communities of users create their own practices and norms for CMC. Therefore, it is difficult to generalize about CMD practices as a whole, or for a single mode of CMC. While studies on CS in CMD tend to focus on interactional aspects of the practice, Dorleijn and Nortier (2009) suggest that a structural analysis of CS in informal CMD texts may shed light on the grammatical issues of the phenomenon. The authors note that a comparison between spoken conversation and written CMD may indicate conventionalized patterns of CS. The discussion in the present chapter is limited to written CMD and focuses most on private, semi-synchronus one-to-one Instant Messaging (IM). A comparison between spoken language and CMD is found in Chapter 6.

4.1 SPOKEN LANGUAGE AND WRITTEN CMD

One comparison that has received ample attention in CMC research is how new written practices compares to spoken language. This type of analysis typically relies on identification of differences in spoken and written communication (Biber 1988, Crystal 2006). Yet it is impossible to make a single claim about the speech-like nature of CMD due to the variety of practices and environments encompassed by this term. A person may write an email to her employer in a manner close to traditional written genres, next chat via IM with close friends in an informal and spontaneous way without editing, and finally write and judiciously edit a blog post. Each of these activities is one mode of CMC, each with its own function and stylistic expectations.

Informal written CMD displays many factors that show users wish to imbue it with more 'speech-like' elements by using 'emoticons' or 'smileys' to convey emotions, tone and basic body language. Capital letters may be used for emphasis or 'shouting.' These traits may be seen in synchronic modes of CMC, such as Instant Messaging or chat forums, as well as in Short Messaging Service (SMS or text) messages due to character limitations. However, Hinrichs's (2006) analysis of email concludes that it is the level of informality and lack of editing that give this impression while a closer examination reveals that written non-synchronous CMD is more like traditional written forms in terms of function. This informality allows speaker-authors to use a colloquial form, such as informal vocabulary, non-standard dialect or CS, in written form.

Despite the great interest in where on the communication spectrum informal CMD language lies, comparisons between spoken language and written CMD of the same speakers is rare. One notable exception is Tagliamonte and Denis's (2008) corpus-based investigation of Instant Messaging (IM) language among teenagers in Toronto. Starting from the strong critiques of IM language in the media, the authors collected IM chat

histories from 71 teenagers and created a spoken corpus from a subset of these speakers. They found important differences between IM and the spoken communication of their participants, in accord with past research that IM is a hybrid between spoken and formal written language but one that is considerably different from either of the traditional forms of communication. Most notable is the juxtaposition of formal and informal elements in the IM texts, as the traditional rules of written language and the expectations of informal spoken exchange can be bent to create a unique style of communication. For example, informal forms like *gonna* occur in the same turn as the formal *must*, despite the fact that the latter is absent from participants' speech. Tagliamonte and Denis conclude that IM as a mode of CMC is similar to both speech and traditional writing while having its own unique conventions.

4.2 INTERNET USAGE IN MOROCCO

The basic hurdle to CMC participation, in any language, is Internet access. While current American college students may not remember a time without Internet, the situation is markedly different in Morocco. In 2001, the earliest Moroccan government figures for Internet usage, just 1.3% of the population had on-line access, whether at home, work or through a public space (ANRT 2001). The early adopters would have lived in urban areas where telecommunications companies had established networks and were likely more educated and of higher socioeconomic status than the general population. A limited sample was analyzed by Berjaoui (2001) in a study of Moroccan chat forums where he found a high rate of Moroccan Arabic (further discussed 4.3.3 below). His conclusion contrasts sharply with Atifi's (2003) findings that Moroccans used primarily French in the Internet forums that he investigated, where Arabic was generally confined to salutations and

cultural references. This could be an indication of differences between CMC modes or between groups that participate in each; the distinction is impossible to make as the metadata for each group is lacking. These early studies represent the only existing literature on Moroccan CMC and can serve as a point of historical comparison with the current data.

The use of the Internet has grown steadily since that time. Today Internet connections are available on smart phones, giving the possibility for users to connect wherever they might be found. The Moroccan Agence National de Réglementation des Télécommunications (ANRT 2014) reported that 56% of Moroccans between 5 and 74 used the Internet in 2013, the year that the data for the current study was collected. These users connect to the Internet in a variety of ways: 84% of Internet users in Morocco access Internet through a mobile phone carrier, either through a phone directly or through a 3G USB modem (Oukarfi 2013). Others may have home computers or utilize Internet cafés, a common connection option in Morocco. At the time of the current study, many young adults used USB modems, a device that can be connected to any computer with a USB port and used anywhere in Morocco that a mobile phone network exists. This form of access can be purchased for set lengths of time, from a single day for 10 MAD (approximately 1.25 USD) or for a 30-day period for 200 MAD (or 25 USD). The flexibility of subscription price and duration makes it possible for more modest households to access the Internet and participate in written informal conversations.

Despite the remaining obstacles, Internet use has created new domains for informal communication in Morocco, including personal emails, Internet forums, and social media such as Facebook and Twitter. However, access does not guarantee ease of usability, particularly for speakers of a nonstandard language variety. When using the Internet, Moroccans must choose which language or languages to employ. The most common choices are Arabic, French and English. For those limited to Internet use via cell phone,

Arabic may not be a straight-forward option as many cell phones in Morocco are not sold with support for Arabic script, and activating such support may require additional technological expertise that is beyond the average user. In addition, those who use Arabic must choose between their local dialect and Standard Arabic. For the former, an additional choice exists between using Arabic and Roman script. However, although Moroccans are educated primarily in Standard Arabic, this does not necessarily translate into familiarity with computer use in Arabic. Young adults learn to use computers in French when formal classes are available.[13] The strong association of French with technology may partially explain why 75% of Morocco's 8.6 million Facebook users are reported to use the website interface in French and only 17% in Arabic (Asharq Al-Awsat 2012).

4.3 WRITING A NON-STANDARD LANGUAGE

Arabic dialects are often devalued and considered unfit for writing. For many speakers of 'non-standard' language varieties like vernacular varieties of Arabic, the Internet provides the first opportunity to write their native variety in informal written communication. In this medium, speakers of varieties without overt prestige may choose to draw on their full linguistic repertoires as communicative resources and incorporate features of informal spoken language in order to express themselves in writing. Writing in a language without an official written form can be at once an act of resistance to the existing standard and a statement of identity.

The case of Arabic is similar to regional dialects of many other linguistic varieties, whose status is often also seen in diglossic relation to the lexifier language. Many of these studies are concerned with the orthographic conventions and speaker adherence to, or

[13] As an anecdote, a Moroccan friend helped me write a description of my research in Arabic. After I saw her hunt and peck typing method, I offered to type and she noted that I type faster than she does in Arabic.

deviation from, 'expert' recommendations for orthography. For this reason, the findings on British Creole and Jamaican Creole can inform the way in which we investigate written Arabic-French CS and suggest motivations for why speakers may use their native variety in writing. At the same time, the social and political settings of these creoles differs substantially from Arabic-French CS due to the prestige associated with each and the similarities, or differences, in alphabets used by each variety. Competition for script choice, *digraphia*, is also faced by speakers of languages that do not use the Roman alphabet (Themistocleous 2010, Androutsopolous 2009, 2013, Lee 2007). For these reasons, the findings on written Creoles are explored first in 4.3.1, followed by a discussion of digraphia in 4.3.2. The current literature on written Arabic dialects is explored in 4.3.3 and ends with a description of the writing system of MCA observed in the corpus.

4.3.1 Writing a non-standard language: Creoles and Pidgins

The writing systems of two varieties of English-lexifier Creoles have been well-described in the literature. Jamaican Creole, the focus of Hinrichs (2004, 2006, Hinrichs and Deuber 2007, Hinrichs and White-Sustaita 2011), and British Creole, investigated by Sebba (1998, 2012). These creoles are similar as British Creole is based in part on Jamaican Creole and used as an identity marker for black/African-descent speakers in England (Sebba 1998). Each creole had been used in poetry and dialogue in novels before the advent of CMC, which may have provided an awareness of how the dialect could be written before large numbers of speakers began using the Internet. One difference between these varieties is that there are efforts to standardize Jamaican Creole, but not British Creole, due to the language situation in each country.

Speakers of Jamaican Creole and British Creole have one orthographic model in Standard English, the lexifier language of both. Creole experts have proposed separate

orthographic systems based on phonology that make to reference to Standard English orthography. Yet speakers diverge from this available norm to distinguish their production from the standard forms. Such divergences can be motivated by semantic distinctions, as with Jamaican *yaad* meaning 'home, house, home country' as opposed to Standard English *yard* or the grassy area near a house (Hinrichs and Deuber 2007). Distinct spelling may also be employed to represent phonetic traits of the creole, such as *de* for *the* or in Creole writings to challenge the standard language forms, as the spelling *blakk* instead of standard *black*. Sebba (1998) argues that British Creole is meant to be a challenge to the conventions of Standard English and by extension to the larger culture with which it is associated. Using English also facilitates reading for those accustomed to Standard English spelling conventions, whether they speak Creole or not.

While connections between meaning or phonetic form and spelling are easily observable in writing, Hinrichs and White-Sustaíta (2011) show that a deeper level of difference exists in Jamaican Creole orthography. They start from Sebba's (1998) assessment that speakers used non-standard orthography to mark their opposition to mainstream expectations and convey opposition to Standard English and the culture that it represents or an 'anti-standard.' Yet not all speakers of Jamaican Creole choose to write the language in the same way. A statistical analysis of orthographic practices indicates that speakers living outside of Jamaica use more Standard English orthography in their writing than those who live in Jamaica. The authors also explore gender and identify a tendency for female writers to use more Standard English spelling conventions, as would be expected in many communities. However, an interaction exists between sex and spelling choice. Female writers use Creole orthography significantly more often when the intended meaning of a lexical item is available only in Jamaican Creole, such as *mi* for Standard

English *my*. While less common overall, female speakers use non-standard spelling to reduce ambiguity in their writing.

This brief description provides several points of comparison for other unwritten dialects. First, it is clear that CMC users do not wait for an 'official' orthographic system, and would likely not follow one were it to be published. Instead, speakers write in a way that aims for clarity in communication, both in semantic terms and in identity construction. Hinrichs and White-Sustaíta's study indicates that intended meaning and sex may each impact orthography, findings that validate inquiry into the effect of these factors in other contexts.

4.3.2 Digraphia

Choosing a script in which to write one's native language is relatively rare. Communities with a long history of literacy practices use or modify the established script with little question. As Sebba notes: "Script selection often takes place at the point where a community begins to practice literacy in its own language, but typically the script adopted is one that is already in use to write another language which up to that point has been the main or only language of literacy" (2007:27). This characterization of script choice may enrich the analysis of Moroccan Arabic for several reasons. First, there is the question of a community's "own" language. While MCA is the native language of the vast majority of Moroccans, many say that their mother tongue is Standard Arabic due to its religious importance and political status. Davies and Bentahila (1989) found that, when asked, Moroccans may identify as their "own" language their first language, the language with which they identify the most, or the language of their grandparents, even if the individual questioned does not speak the same language. Identifying a single language at the

community level may be easier in most areas of Morocco where the majority of daily interactions are carried out in MCA.

Second, although some artistic representations of language used Moroccan Arabic before the spread of the Internet, most MCA speakers did not have a format in which writing their native dialect would be appropriate. The informal communication on the Internet may have been the first opportunity for a wide swath of Moroccan society to begin creating a literacy practice in their community language. The reasons behind the current ambivalence of Moroccan Internet users regarding script may also be hinted at by Sebba. Both French and Arabic have been used for literacy in the last 50 years, both included in compulsory education, both present in written press. Abbassi's (1977) respondents reported a preference for newspapers in French, which indicates that the generation he studied may have had a stronger affinity for literacy in French. Standard Arabic, as the language of the Qur'an, is often seen as the ancestor of MCA, but its religious and cultural importance may lead speakers to distance their own dialect from it. Literacy practices and preferences are not the only aspect of script selection in digraphic situations. Other recent situations of script selection given by Sebba are from former countries of the USSR that rejected the Cyrillic alphabet; thus selection of a new script was part of establishing their independence and cutting political and social ties.

There is also the question of technological limitations with regard to CMC writing. Early Internet was limited to ASCII encoding, a type of encoding that allows 128 characters, including upper and lower case of the most common letters in European languages, numbers, and common symbols. Use of ASCII characters may have become entrenched among Internet users, leading them to associate Roman script with the Internet in a broader sense. Themistocleous (2010) notes that this is still the only encoding on certain parts of the Internet, such as Internet Relay Chat (IRC). In this way, encoding

limitations in that mode of CMC impose the Roman alphabet on the speakers of Cypriot Greek. As a result, her entire corpus of Cypriot Greek chat forum conversations is in the Roman alphabet, despite the fact that some Cypriot Greek speakers would write their dialect in the Greek alphabet in other online contexts where there is better support for other types of character encoding. Similar to Arabic dialects, users may choose the spelling that they find best reflects the local pronunciation, meaning that readers cannot rely on word recognition and must sound out each word as they read. While some (Ennaji 2005, Asharq Al-Awsat 2012) report that keyboards with Arabic script remain rare in Morocco, all keyboards encountered in Meknes during the period of this research project included Arabic and Roman characters, whether at Internet cafés, the university, or for personal use.

A broader question in the use of Roman script by speakers of unwritten varieties is whether the practice will spread to off-line settings. Palfreyman and El Khalil (2003) found that 25% of their university-aged respondents reported having used or seen RA off-line as a sort of 'secret code' in communications in order to not be understood by other individuals such as teachers, test proctors, and parents. This practice has not been reported, anecdotally or otherwise, in Morocco. Worried about broader use of MCA in RA, Berjaoui stresses that the use of the "French" alphabet may create a cultural dependence and warns that, were MCA to become standardized, the selected script would also become a national symbol in part. A danger specific to minority communities is that such a shift in script will lead to a language shift from the home language to that of the host community (Angermeyer 2012). Such a possibility seems remote in Morocco where MCA is the majority language, but it might negatively impact the use of Standard Arabic in print media were its use to spread.

Association with French is not the only reason to use RA. Angermeyer (2012) points out that the power of English as an international language is another vehicle for the spread of Roman script. As Moroccans are increasingly aware of the utility of English on

the international stage, this may reinforce the practice of writing MCA in RA. Whether the connection with French or English is stronger, or becomes magnified by both, the decision to write MCA in RA instead of Arabic script is an act of departure from the literary heritage of Standard Arabic as an official language and the language of the Qur'an, rendering such a choice laden with social, political and religious implications.

4.3.3 Written Arabic dialects

It is typical for native speakers of Arabic dialects to consider only SA a full 'language,' referring to their own first languages as 'dialects' that they generally do not view as worthwhile in writing (Ennaji 2005, Sayahi 2007, Benrabah 2007). Yet MCA was used in written media from references to magazines in the dialect from the 1980s and 1990s (Ennaji 2005, Brustad p.c.). Such print sources are no longer available, and no online sources seem to have taken their place despite the popularity of a number of online news sites are dedicated to news within and related to Morocco. This is not the case of all dialects, evidenced by the more than 13,400 articles on Wikipedia written in Egyptian Arabic at the time of writing (Wikipedia 2015). Plays, novels and advertisements may use the local dialect even in countries where the dialect is not championed in writing, in order to be understood by a broader audience, evoke the social values of the dialect, or stress the fact that they are local. Today MCA in RA is used in a limited way in printed publicity and is generally restricted to symbolic portions of advertising related to cell phones, as in Figure 4.1.

Figure 4.1 Mobile telephone service advertisement

(Top left in RA) "Free Hour" (Bottom left in Arabic script) "with everything" (Right, in French) Prices, benefits, subscription information

Still, no dialect has a standardized writing system, meaning that transcriptions are often ad hoc, whether they be in the Arabic alphabet or in the Roman. This underscores the lack of a community of practice by writers of MCA, supported by language ideologies that emphasize a distinction between SA and the dialects, with SA seen the only 'correct' Arabic variety for formal writing.

One difficulty of typing in Arabic dialects in Arabic script stems from the phonetic differences between dialects and SA. Pronunciation differs greatly between certain dialects, and a single phoneme may be present in two dialects, but represented in SA by two distinct letters. Berjaoui (2001) points out that a country's 'dialect,' though often called by a single name, is actually a collection of regional dialects that may vary substantially. For example, the verb قال "he said" has the phonetic form /qal/ in SA and certain regions of Morocco, but is more commonly realized as /gal/ with the /g/ phoneme one feature of Moroccan Arabic. In addition, <ج>, generally pronounced /ʒ/, may also become /g/, as in the common work /gləs/ ("sit!") Both <ق> and <ج> are pronounced as in SA in other lexical items, leaving speakers the choice to change the spelling of قال, and thereby reflect MCA pronunciation, or to remain faithful to the SA spelling and by extension imply a standard

pronunciation. Today /g/ is often written <ك>, with or without a diacritic, although this letter is otherwise pronounced /k/. This practice may be influenced by the use of this letter in Arabic spellings of cities such as Agadir (أكادير) and other words of Berber origin as an adoption of existing practices.

The insistence on the differences between MCA and SA phonology may also have an ideological aspect. Using a phonemic inventory based on multiple descriptions of MCA, Berjaoui (2001) gives 16 MCA phonemes that are not found in SA, which he concludes suggest the extent of the phonological differences between the dialect and standard language. However, any student in Phonology 101 would be skeptical of his list due to the variety of sources from which his list was created without question and the lack of minimal pairs or other examples cited. The possibility of feature spreading is also not investigated despite the fact that thirteen of this sixteen proposed MCA phonemes are pharyngealized forms of existing MCA phonemes; furthermore, the examples in which most of these phonemes are given have only pharyngealized consonants. His own evidence therefore calls his conclusion into question on empirical grounds, suggesting that attitudes and ideology play an important role in his description of the dialect. The same attitudes and ideology may lead speakers today to separate the Arabic varieties by script in their writing practices.

Early Arabic-speaking technology users had no choice in alphabet; Arabic script was not available on most computers in the early years of the Internet and even today right-to-left text support is lacking in some word processing programs, Internet interfaces, and email clients. Without support for Arabic script, text either had to be in a foreign language, typically a European language, or in Arabic written in the Roman alphabet, sometimes called ‘3arabizi,’ ‘Arabizi,’ or ‘Arabish,’ as combinations of the words *Arabic* and *English* in English or /ʕrabi/ and /inʒlizi/ or /inglizi/ in Arabic, or ‘Romanized/Latinized Arabic’

(Palfreyman and Al Khalil 2003, Warschauer et al. 2006, Yaghan 2008, Bjornsson 2010, Abu Elhija 2014). Romanized Arabic (RA) will be used here because it does not imply a specific contact language, given that it is French, and not English, that has the greatest influence on Moroccan RA. When writing in RA, Arabic speakers tend to relate the sounds of their dialect to the expected spellings of that sound in the European language. In countries where English is the predominant European language, such as Egypt, speakers often spell /ʃ/, Arabic <ش>, as <sh> while <ch> is the typical spelling in French-dominant countries, including Morocco.

The earliest study on CMC in MCA was carried out by Berjaoui (2001) who examined the orthography of MCA as utilized in IRC chat rooms. He found a high use of the Arabic alphabet as well as RA, with a low level of variation in orthography for each alphabet. He does not mention use of French within the forums that he investigated.[14] Around the same time, Atifi (2003) found that users preferred French on the forums she investigated. Again, these participants would have been early adopters who likely had a higher level of education and socioeconomic status than the general population. Without ethnographic information on the participants, no certain conclusions can be made, but it seems that the forums from each study had their own expectations and may have been popular with distinct user groups. His final statement on the topic is:

> Until some official device for the MA orthography has been devised, the most widely used means for the writing of this dialect, which guarantees written and distant communication in Morocco (together with both cultural and symbolic loads) seems to be the borrowed orthography of SA. (p. 463)

As can be found from quick consideration of Moroccan newspaper comments, forums or Facebook pages, today no alphabet can confidently be considered 'the most widely used'

[14] *Forums* is the plural used here on the topic of modern internet use as distinct from the *fora* of a Greek city state.

as MCA is commonly found in both the Arabic and Roman alphabets. As described further in Chapters 5 and 6, young adults seem to be most comfortable with MCA in roman script.

Despite vast differences between the use of Internet in Morocco at the time of his study and today, Berjaoui's conclusions hint at what we might find in written MCA today. First, he finds that no existing solution is both easy for users and accurate in representing MCA as both the Arabic and Roman alphabets lack certain phonemes present in the dialect. He notes that a phonetic transcription would negate these difficulties, but that such a system is not realistic for a naïve user – similar to the situations of creoles described by Sebba and Hinrichs. Second, there is a great deal of orthographic variation when MCA is represented in either script due to what the author perceives as a lack of knowledge by naïve speakers; this may be more objectively characterized as the difficulty of creating phoneme–symbol correspondences in either script as Roman and Arabic characters do not easily lend themselves to Moroccan phonemes. Finally, Berjaoui's description itself provides a point of comparison between the use of MCA in RA today as well as in other dialects. Recent descriptions of written RA can be found for Egyptian (Bjornsson 2010) and Levantine Arabic (Abu Elhij'a 2012). These descriptions indicate that written norms are emerging, but that a high degree of variation remains between users. Intra-speaker variation is also visible in the written dialects as a user may choose different representations of a single phoneme across multiple tokens of the same lexical item. The main emphasis of this study is not orthography, as Bjornsson and Abu Elhij'a's theses demonstrate that this topic merits in-depth discussion for each dialect.

Today support for Arabic script is commonly available on computers, although mobile telephones are still sold in Morocco without support for Arabic script, reinforcing the use of RA even if individuals might use Arabic script in other media. As noted in 4.1 above, many Moroccan Internet users connect through cell phones, making this an

important aspect of script choice. Asked if their cell phones have support for Arabic script, 64% of participants in the current study report that theirs did not. Certain participants mentioned that they checked their phones for this function while completing the survey – an indication that the ability to use Arabic on a cell phone is not a factor in choosing a device. The difference in character encoding necessary to send Arabic script also 'costs' more; each character takes up more digital space than the unaccented Roman alphabet, effectively cutting the standard text message length from 140 characters down to 60 on most carriers. An additional possibility offered by Moroccan mobile networks is a basic Facebook client usable without a smartphone. Regular mobile phones are less likely to have Arabic support or an Internet connection. While the number of users taking advantage of this method of access is unavailable, its existence may be another motivation for users to write in RA.

If technical limitations were the only reason to use RA, as Aboelezz (2009) notes, we should be witnessing a decrease in the use of RA. Recent research on RA in several countries concur that script choice has an array of other implications. Aboelezz (2009) states that dialects are 'almost always' written in RA, which she suggests may empower the dialects. Bianchi's (2012) data point to the importance of topic in script choice in Jordan as forums related to hobbies, gender or age group interests, family or friends, and work or study contain more Arabic in RA than in Arabic script, while forums related to humor, poetry, and most local culture topics were primarily in Arabic script.

Warschauer et al.'s (2006) study of 'young professionals' in Egypt presents multiple aspects of script choice in an Arabic dialect. They assessed language use in three CMC modes (formal email, informal email, online chat) via self-report surveys on Internet usage and in-person interviews of select survey respondents. They found that English was the most common language in all contexts, but Egyptian in RA was used by nearly half of

the respondents in informal email and online chat. Few respondents use Egyptian in Arabic script. Their results also suggest a connection between use of RA and code-switching. About half report using English and Arabic in informal email as well as online chat. While the motivations of language choice are not fully explored within their article, certain participants report code-switching into Egyptian Arabic when they cannot think of a word in English.

As speakers have different motivations for using RA, research has also considered what types of speakers may choose to employ it. Zoabi (2012) examined Facebook status updates and wall posts from nineteen different Arabic speaking countries and found that the data from Morocco, Tunisia, Algeria and Lebanon was all in RA. At the other end of the spectrum five countries used only Arabic script, while the data from the other ten countries contained both scripts with a minimum of 48.6% wall posts in Arabic script used in Egypt. For this reason she speculates that an aspect of the French colonial system has affected script choice in these countries. Aboelezz (2009) found a connection between RA and code-switching in her Egyptian group email data: RA was used most often in emails that also contained English. She found both practices to be more common in the mailing group of Arabic speakers proficient in English than in the group of mixed English ability.

The situation of Warschauer et al.'s respondents in Egypt may apply to many aspects of Moroccan Internet users. First, the Egyptian respondents report not being proficient typists in Arabic as they learned to use computers in an English environment. In Morocco, computer literacy courses are taught in French, which gives students experience typing in the Roman alphabet, but not in Arabic. Second, respondents reported that the computer interface at their places of employment were in English or that their bosses often use English with them through written CMD. The same can certainly be said of French in Morocco. Finally, the authors found that two factors correlated significantly with use of

RA: years of Internet use and careers in information technology, a field associated with English. It may be that these individuals began using computers and the Internet in English or before sufficient Arabic script support and have become comfortable with the RA writing system.

Several studies note positive and negative aspects of using RA. Before beginning the study, anecdotal stories from young Moroccans indicated that some felt that the dialect written in Arabic script 'felt wrong' and should be reserved for SA as the language of the Quran, while others found any Arabic in RA 'confusing.' This clear division among individuals of a similar age range validates the investigation. It may be that some speakers prefer to visually separate MCA from SA through the use of RA, while those unfamiliar with the RA writing system avoid it through use of French or MCA in Arabic script. Warschauer et al. (2006) posit that the informality of CMD and lack of overt rules may encourage RA. At the same time, they warn that when speakers choose to use RA, the domain of SA may shrink as it is pushed from the informal written domain from RA on one end and by English (and French) as the dominant languages of the Internet on the other.

While a surprisingly large body of work explores RA, it generally centers around the orthographic conventions of the written dialect and the correspondences between written form and phonological features of each dialect (Palfreyman and El Khalil 2003, Warschauer, El Said and Zohry 2006, Yaghan 2008, Bjornsson 2010, Abu Elhija 2012). The connection between RA and bilingual ability (Aboelezz 2009) and setting or technology experience (Warschauer et al. 2006) signify possible motivations for script choice and explain differences between speaker groups. These lines of research differ greatly from that of the more common interactional analysis found for CMD of other language pairs, discussed above.

4.3.4 A brief description of Romanized Moroccan Colloquial Arabic orthography

A basic description of Romanized MCA (RMCA) orthography is given here to help the reader understand the given examples of written language as produced by participants. I emphasize that this is not a full account of the current trends in writing MCA, which merits a full treatment of its own. Arabic written in Roman script is often considered a type of transliteration system. However, RMCA should be considered a writing system in its own right, and not a transliteration system, as transliteration implies use of a script different from the typical script of a given language. No standardized written form of MCA exists; while many of the sounds in MCA are shared with Standard Arabic and its script, others are absent from it.

A great deal of variation exists in RMCA, as visible in the common phrase أن شاء الله generally pronounced in MCA /nʃalla/, 'God willing.' Eleven unique spellings of this phrase are found among the twelve participants who use it in the IM chat, seen in (4.1)

4.1)

inchaalah	P12032
Inchaalah	P11292
inchaelah	P12232
inchallah (3 times)	P11021
insha allah	P11291
inshaalah (2 times)	P11051
inshaelah	P11202
ncha2llah (2 times)	P10282
nchaellah (2 times)	P11201
nchalah (2 times)	P10281
nchalah	P11271
nchallah	P12231

Several important commonalities stand out in the eleven spellings found in (4.1). First, ten of the spellings represent the phrase as a single word, as it is spoken in MCA. P11291 breaks the phrase into two words, but not the three that it is in Standard Arabic. It may be that he uses a space to represent the presence of the glottal stop, /ʔ/, in Standard Arabic. One participant, P10282, clearly represents the sound /ʔ/, <2>, although this phoneme is not part of the MCA pronunciation. P10282 is also the only Arabic instructor in the group, which may have influenced her addition of the SA /ʔ/ phoneme. At the same time, P10282 is among those who begin the word directly with <n>, which is closer to the MCA pronunciation due to the fact that leading vowels are systematically dropped in this dialect. Many trends are visible in 4.1 and it is included here to highlight the interspeaker variability of RMCA. Geminates are rare in RMCA, despite the fact that they are part of spoken MCA, but are used in this phrase by 6 speakers. The phoneme /ʃ/ is typically represented as <ch> due to French influence, but 3 speakers use <sh>, similar to English. While two of these participants are English majors at university, high proficiency in English cannot be the only reason to use this orthography as P11051 has much less exposure to English, which she learned through studying at a language center a few hours each week.

Despite the level of variation, there are clear general patterns in RMCA, given in Table 4.1. The sound-symbol associations used by speakers are largely phonetic, although the pharengealized phonemes share a symbol with their non-pharengealized counterparts, such as sounds are leveled with their non-representation of the dialect as seen in (4.2)

4.2)	llah	i3tina	sber	w	safi	(P11282, written)
	llah	yʕṭina	ṣbər	w	ṣafi	re-transcribed
	God	give=us	patience	and	that's.all	
	God give us patience, that's all we can do					

This leveling may be due to the difficulty to distinguish between the two through the Roman alphabet.

A very different departure from the phonetic form is seen in the occasional addition of <e> to the end of certain words when the symbol does not correspond to any phonological content, such as in (4.3)

(4.3) amatkhafche (P12032, written)

a	ma	t=khaf	che	separated into analyzable parts
a	ma	txaf	š	re-transcribed
oh	NEG	2nd.PRES=fear	NEG	

oh don't be scared

Final silent vowels are common in French and English, but seem to serve no purpose in RMCA. The unnecessary nature of such silent vowels is evidenced by their rare occurrence, which most participants avoid.

Table 4.1 Sound-symbol correspondences in Romanized Moroccan Colloquial Arabic

RMCA	Transcription[15]	IPA	Arabic script
2*	ʔ	ʔ	ء
a	a	a	ا
b	b	b	ب
t	t	t	ت
--	--	θ	ث
j	j	ʒ	ج
7 / h*	ħ	ħ	ح
kh	x	x	خ
d	d	d	د
--	--	ð	ذ
s	s	s	س
ch / sh	š	ʃ	ش
s	ṣ	sˤ	ص
d	ḍ	dˤ	ض
t	ṭ	tˤ	ط
--	--	ðˤ	ظ
3	ʕ	ʕ	ع
gh	ɣ	ɣ	غ
f	f	f	ف
9 / q*	q	q	ق
k	k	k	ك
l	l	l	ل
m	m	m	م
n	n	n	ن
h	h	h	ه
w / o / ou	o	o / u	و
i / y / ey*	i	i	ي
e	ə	ə	--
* Rare			

[15] Cf. Figure 1 in §1.3

4.4 WRITTEN CMD AND CS

Written CS as found in CMD has been explored from a variety of frameworks and with many different goals. However, the available literature on written CS confronts several difficulties as there is no framework for studying written CS (Sebba 2012) and therefore many studies utilize frameworks created for spoken communication, such as Auer's Conversation Analysis model (1984, 1995) and Myers-Scotton's Markedness Model (1993). Recent studies have focused on CS as present in a variety of types of CMC. The past findings that inform the current approach are considered in this section.

In response to the lack of framework for written CMD and particularly CS in CMD, several authors have identified aspects of these practices that separate them from traditional spoken CS and the spoken modality. Sebba (2012) recommends five facets for a sound framework of written CS: interactivity, synchronicity, sequentiality, permanence, and visual aspects of the text. Kytölä (2012) notes that in order to investigate CMD in a complete, replicable way, multiple methods should be used. He recommends corpus-linguistic methods combined with ethnographic information on participants or similar users when the exact participants may be unknown or unreachable. He also stresses the importance of qualitative analysis of pragmatic and interactional functions of CS, the main focus of CS CMD studies to this point. Androutsopoulos (2013) highlights potential differences between private/public, one-to-one, one-to-many interactions. Private CMD is often between speakers who have common background knowledge and thus may have common CMD practices and shared linguistic repertoires.

One potential way to analyze CS in CMD is by comparing it with that of oral speech. While pragmatic and interactional functions of CS are important and represented in the literature, another aspect of the phenomenon is largely absent: structural properties. Dorleijn and Nortier (2009) note that such an analysis could improve the current

understanding of the differences between CS practices both by modality and by community. As the structure of CS has been of primary interest in research on the spoken phenomenon, the authors recommend considering the same factor in its written correlate. The act of writing requires reflection on issues such as spelling and grammar that are irrelevant or easy to ignore when speaking, even in synchronous chats, which are typically not edited (Lee 2007). Dorleijn and Nortier note that little work has been completed to identify the effect that this greater attention has on CS and its structure. While not structural, Deumert and Masinyana (2008) suggest that the English-isiXhosa mixing employed in text messages (SMS) by their participants mirrors the participants' unmarked vernacular, citing that 23% of SMSs are mixed. It may be that CS in CMD is similar to that in spoken language because of its function within discourse and the dialogic context (Androutsopoulos 2013).

Harris (2010) carried out a comparative quantitative study of CS in speech and written CMD. He examined blogs from Quebec and France, a type of one-to-many, one-way CMD, in addition to speech in reality television shows from each country. As both of these are informal contexts, it might be expected for speakers to use similar communication styles and CS practices in each. Like Tagliamonte and Denis (2008), he found noteworthy differences between written and spoken modalities. At the same time, an important distinction is also uncovered; Anglicisms are more frequent in his spoken Quebec French data than in the written, while the reverse is true of the French spoken and written data from France. This result supports the need for considering the use of CMD and CS within it as a community-specific practice such that generalizations cannot be made for CMD as a whole. Speakers choose how to employ this new form of literacy and those from different communities may do so in contrasting ways.

Another use of CMD in a former French colony is described by Lexander (2012). She examined the language of Short Message Service (SMS) messages among young Senegalese, selecting this type of CMC due to its popularity among speakers. Speakers of multiple language backgrounds included French in their texts, and sometimes in predictable ways. For example, CS was sometimes flagged by capital letters, or found in what she calls 'ring switching,' the use of one language for the start and end of a message with the other for the body. Relating to communication modality, she noted that speakers of Wolof and French often used more French in SMS than in their speech. In addition, participants do not use standard Wolof orthography in their SMS messages, even those who they report doing so. This confirms the general finding that written CMD has its own conventions and must be considered apart from traditional spoken conversation and formal writing. In addition, the use of capitalization indicates that speakers are aware of the nonstandard language use and find it appropriate to flag this even in an informal medium.

The final example relates to script choice and the use of Instant Messaging (IM). Lee (2007) used a combination of IM logbooks and interviews to gain a broad understanding of the use of IM by her speakers. Most of the conversations took place with friends, who preferred to use Cantonese in order to reflect their typical offline spoken interactions. English words were often inserted at times where they are used in spoken conversations. At the same time, certain technical constraints increased the use of English. Certain words were used in English when the Cantonese characters were difficult or slow to type in order to increase speed of responses. As expected in semi-synchronous CMD, her participants reported minimal editing and a lack of attention to grammar and spelling; comprehension and speed were the primary goals due to the immediate back-and-forth nature of IM conversations. In fact, participants note that speed is an important factor in word choice at times because they report using less English in SMS due to the greater

amount of time to respond. In an earlier analysis of email and ICQ ('short' for the sentence 'I seek you') data, the first Instant Messaging client, Lee (2002) found this to be true; ICQ exchanges contained CS more often than emails. The conscious emphasis on speed in these exchanges, sometimes at the expense of correct grammar or spelling, demonstrates that speakers are more concerned with the content of the message than its form and utilize whichever form will achieve their communicative goals.

4.5 CONCLUSION

The findings of previous studies highlight key questions in Arabic/French CS in CMD. IM is expected to be a unique medium that shares aspects of traditional spoken conversation and traditional writing. The immediate focus is on the structure of CS as analysis of this aspect of language use may expose conventionalized uses of French and MCA that are shared with spoken conversation as well as those that are unique to written CMD. In this way, the current aim is similar to that of Tagliamonte and Denis: to understand the extent to which written informal CMD has become an independent communication modality. The current study utilizes Instant Messaging alongside spoken conversation (see Chapter 5 for further detail) by the same speakers in order to maximize comparability of the data. As one of many types of written CMD, IM can be defined as an interactive, one-on-one form of CMD that is private, synchronous, sequential, and generally temporary[16]. As Androutsopoulos (2013) notes, research on private, dyadic conversations are the least represented in research; the current study will help to fill this gap. This study also contributes to the broader field of CMC as it is the first to address informal written Arabic-French CS and the first to utilize spoken and written CS production

[16] Some IM clients delete chat conversations once they are closed while others keep the conversation for a short time, or indefinitely. The complexity of accessing past chats is similarly client-dependent.

from the same speaker pairs to better understand the similarities and differences between structural properties of these modalities.

5. The Study

5.1 RESEARCH QUESTIONS

Following the above discussion, four questions will be explored in this study:

Question 1. What proportion of informal language use by young Moroccans is constituted by French lexical items?

Question 1b. What is the structure of the French lexical items in MCA?

Hypothesis: It is expected that Moroccans will use a low level of French lexical items in MCA-French CS. As there are no published studies that quantify the rate of CS as a proportion of all speech, it is impossible to give an expected level of CS. The structural expectations are that there will be a high proportion of nominal structures, as noted by Bentahila and Davies (1995) in the younger generation and by Sayahi (2011) in all speakers. However, the lack of recent data on the exact location of switches, combined with Ziamari's (2008, 2009) consistent assertion of evolution in MA-French CS, more precise syntactic structure or the relative amount of full Determiner Phrases or Bare Nouns are impossible to predict.

Question 2. Does communication modality (written or spoken) affect the rate or structure of French in MCA-French CS?

Hypothesis: The lack of prior studies precludes strong hypotheses. Due to differences found in typical monolingual spoken and written communication, it is expected that distinctions between the modes will also be identifiable in CS, but how these might be observed is unknown.

Question 3. How do external factors of SEX, FRENCH PROFICIENCY or LANGUAGE ATTITUDES contribute to the variation observed in rates and structure of CS?[17]

[17] These terms are given in small caps when used as extra-linguistic factors and in standard type face when used to refer to characteristics of a given speaker or speakers.

Hypothesis: Based on recent findings, it is expected that rate of CS will be the same between genders. Higher French proficiency is expected to correlate with higher rates of CS. Positive attitudes toward French, as a measure of approval of French in general and CS between MCA and French, are also expected to correlate positively with use of CS. Based on the findings of Sayahi (2011a) and Bentahila and Davies (1995), it is expected that increased French proficiency will be predictive of intrasentential mixing with a greater diversity of syntactic junctures. Sayahi's results also indicate different trends for type of constituent used, with females more likely to use inter-sentential switches or full clauses in French, but more likely to use bare nouns. As language attitudes have not been investigated in connection to constituents switched, it is impossible to make predictions in this area.

5.2 TASKS

Two naturalistic production tasks provide the corpus for this study, paired with an ethnographic background questionnaire and a written French grammar test. A written chat task was completed first via internet (IM task), followed by a spoken conversation. Both communicative tasks were completed by pairs of speakers and all pairs were consistent across the tasks in order to maintain the same relationship between participants, thus ensuring comparability of the written and spoken data. Participants were asked to speak with friends for the production tasks, instead of family members, as young Moroccans reported the highest use of French and CS with friends (Lawson and Sachdev 2000, Chakrani 2010). Despite this direction, two pairs (one male, one female) were siblings.

After the production tasks, participants completed a background questionnaire on their typical language use as well as educational history. The questionnaire included

various aspects of French use, including typical use by family members within the home, city of origin of the participant and her parents, type of schooling. The primary questions for the current analysis pertain to gender, French proficiency and language attitudes. Finally, a short French grammar test was administered, found in Appendix B. I was also present while participants completed the questionnaire and grammar test in order to answer any questions and prevent participants from discussing answers to the grammar test.

Throughout our interactions, I spoke with all participants in Moroccan Arabic as participants had varying levels of proficiency in French and English. Use of the dialect also demonstrated its acceptability in the context of the study and my comfort with the dialect. To further put participants at ease, all participants were informed that a native speaker of Moroccan would assist the researcher with comprehension of conversations if necessary. Certain participants chose to respond to me in French or English. In these cases, I replied in Moroccan Arabic. If a participant pair asked a clarification question a second time in a European language, the same European language was used in response. Regardless of individual use of a European language, it is expected that interactions with a foreign researcher likely impacted speakers' decisions to participate in the study, as some openly questioned me regarding why an American would want to record them or discussed it during the production tasks. My background may have subconsciously impacted speech style as well, prompting a greater use of French. However, it is expected that this factor impacted all groups equally as I interacted with all of the participants.

5.2.1 Written chat (IM) task

The use of written internet-based chat, or Computer Mediated Communication (CMC) in this study is significant in multiple ways. First, it adds to the growing body of literature on written Arabic dialects (Bjornson 2010, Yaghan 2008, Abu Elhija 2012, 2014,

Berjaoui 2001, Atifi 2003). Second, it can be viewed as an example of language standardization, or lack of it, coming from the users in the absence of 'expert' or government guidance (Sebba 1998b, 2012, Hinrichs 2006, Deuber & Hinrichs 2007, Hinrichs and White-Sustaíta 2011). Third, a thorough comparison of the CS found in written CMC and spoken language of the same individuals has not been published but such a comparison may provide evidence of how communication mode and the higher level of planning provided by written CS may affect CS practices. Thus the study has important implications within Arabic, across non-standard languages, and within the field of CS as a whole.

In this task, the participants each connected to the internet via computer and chatted with each other for an hour using Facebook's chat function within the web-based email. The chat is an Instant Messaging (IM) client embedded in the Facebook website. Users can see whether the other user is composing a message, but only see the words written when the other individual presses the Return or Enter key. For this reason, Facebook chat is considered a type of semi-synchronous IM. Facebook's chat interface was chosen due to its popularity with the target population, and validated by the fact that all participants already had a Facebook account. This makes it likely that participants were familiar with the functionality of the chat feature and were comfortable using it. Another benefit of this medium is that Facebook saves the entire chat history between individuals, assuring that the data would not be lost if the internet connection were disrupted, a common occurrence in Morocco, or if participants forgot to forward the conversation immediately to the researcher. All of the computers used in this task were chosen by the participants in private homes or internet cafés based on personal preference and availability. The researcher's computer was specifically not used as it has an American keyboard with a QWERTY layout and no Arabic letters, while Moroccan keyboards use the French AZERTY layout

with the Arabic letters printed permanently onto the keys. Participants were not asked about the style of keyboard used, but it is assumed that a standard Moroccan keyboard was used.

In order to avoid the potential distractions of multiple conversations, participants were instructed to chat only with their partner in the study during the hour of written chat. To facilitate this, participants were reminded of how to make their Facebook chat open only to specific friends. While certain participants refer to communication with other individuals either physically present in the same space or in contact with them through the internet while completing the IM task, these separate interactions appear to have been brief. During the written chat participants were located in separate rooms or separate buildings whenever possible to increase the naturalness of using chat and prevent speaking. However, this was not achieved in all situations; three pairs of participants chatted while in the same room, although reported not speaking to each other, and one pair chatted from different rooms of the same private residence.

Participants were directed to chat about anything they liked for an entire hour. They were explicitly told to use any language(s) they desired and any alphabet(s). Many asked what kind of topics they should discuss, or if they should make up something. In these cases it was stressed to chat normally about everyday topics. Studies, family, friends, sports and gossip were specifically given as examples when the participants remained unsure or doubtful of being able to think of a topic. Some participants seemed to be unaccustomed to chatting for a full hour, as evidenced by prompting each other to suggest a topic at certain points.

5.2.2 Face-to-face conversation

The second task, a spontaneous spoken conversation, is the typical source of data for code-switching research (Bentahila 1983, Bentahila & Davies 1995, Ziamari 2008,

2009, Poplack 1980, and many others.) This data type is key to the study as code-switching remains a primarily spoken phenomenon; the perspective of this dissertation is that the comparison of the syntax in this mode with that of the written IM data broadens our understanding of intra-speaker variation.

The face-to-face task asked participants to sit together in a quiet room to record a spoken conversation for an hour. The conversations were recorded using a Tascam DR-40 digital recorder at 44.1 mHz with a Shure MX 392/S cardioid condenser microphone. The conversation setting varies due to participant choice; eight conversations were recorded in private residences, ten in a university office, and the final three in a university classroom. Due to the presence of other individuals in the house or building, occasional input from non-participants were edited out of the saved recordings. In these cases, any responses from the participants to third parties was retained despite the fact that the speech patterns are likely affected by the change in intended addressee (Bell 1984).

The researcher was present at the beginning and end of all recordings to obtain consent for voice recording, set up equipment, and start and stop the recording equipment. Moroccan Arabic was again spoken with participants, resorting to a European language only if the same question was asked twice. Certain participants were excited to speak French or English with a foreigner and did speak more with the researcher before agreeing to participate or after recording. Conversations directly before recording were kept as brief as possible while being sensitive to cultural norms of politeness. In the task, participants were asked to speak about any topic of their choosing, again for an hour, and told that the researcher would return to the room at the end of the hour to stop the recording and administer the background questionnaire. Once more the use of all languages was explicitly welcomed. If participants expressed difficulty at thinking of a topic, studies, family, friends, sports and gossip were again suggested.

As is well known among social science researchers, recording participants aware of the recording equipment creates the 'Observer's Paradox.' In fact, one participant after hearing the basic research tasks told me, "but you can't tell people you will record them! There are lots of things we will not say, topics we will not discuss if we know someone else is listening!" The strong awareness of this fact by a naïve speaker indicates that the recorded conversations may not be fully representative of everyday speech, but ethical considerations prevent recording speakers who are unaware of the recording. While it is likely that participants adjusted their conversation topics, the experimental process should have minimal impact on the structure of participants' speech.

5.2.3 Background questionnaire

The background questionnaire, available in Appendix A, was presented in a paper and pen format as internet connections could not be guaranteed at recording locations. In addition, use of a computer survey assumes comfort with filling out a form via computer, which may not be true for all participants, and may have prompted the use of more 'standard' language forms. The questions were presented in both French and Standard Arabic on the same paper. Since French is written left to right and Arabic right to left, the questionnaire was in a two column format with the text in French beginning on the left side of the page and the Arabic on the right to allow each language its natural flow and not give a visual preference to either. Writing space was provided in the middle of the two languages, again allowing participants to choose the language of their response without suggesting directionality. Participants were directed to answer in whatever language or languages they wished, and the language of the answers was considered in the analysis.

The first section of the questionnaire focuses on participants' formal education, followed by the languages spoken by their parents and those used in their homes. The fourth

section regarded language use on the internet. The next section contains a language self-assessment in French and Standard Arabic based on function, and finally five questions on language attitudes. In this way it was possible to gain a general idea of the participants' linguistic background from formal education as well as home environment. The current analysis focuses on French proficiency, gender, and language attitudes of participants.

The importance of social class on language use in Morocco is visible to visitors and discussed by Moroccans as well, but is difficult to measure and for this reason was not included in the current analysis. As Chakrani (2010) notes, young adults are unlikely to be able to give the approximate salary of their parents either due to lack of knowledge or social pressures. His solution was to ask the profession of the father as this was found to be reliable by Wagner (1993). Once the professions in question were identified, Chakrani asked a separate group of Moroccans to estimate the yearly salary for each profession. For each profession, he placed the averages of these estimates on a salary scale for Morocco based on information from the High Commission for Planning (2009) and the World Bank (2009) in order to objectively place participants into social classes. This methodology was intended in the current study, however difficulty was encountered in the assignment of social class by profession. For example the most commonly reported parental profession was استاذ which can be the American English equivalent of 'tutor,' 'teacher' or 'professor,' and the social class associated with this range of work can vary greatly. The same term in MCA may refer to an individual who does part-time tutoring, and is likely of lower social class, an employee of a language center for foreigners, whose salaries may vary greatly, a public education teacher, or to a university-level professor who also teaches private lessons during the summer and thus achieves a high salary. A similar difficulty is found for other professions. Another difficulty was noted from the two sibling pairs and their descriptions of their father's work, one put 'entrepreneur,' which is typically rated as an upper-class

profession, while the other described him as a 'construction worker' which is typically lower-class. These difficulties lead to excluding class from the current analysis despite its apparent importance in language use in Morocco.

5.2.4 French Proficiency evaluations

Noted in §3.24 above, fluency in French may also impacts speakers' use of CS and its structure. For this reason, two measures of French Proficiency were included in the written portion of the data collection, first a task-based self-assessment of language proficiency and second a short written evaluation of French grammar was conducted as the final task of the experiment. The self-assessment was completed as part of the Background Questionnaire as found in Appendix A. This assessment is similar to that used by Anderson (2006). Its inclusion as a measure of French proficiency provides a subjective measure, as the common use of French on television or with foreigners in Morocco may give students a higher functional ability in the language than indicated by the results of the written grammar test.

A shortened version of the Oxford French Placement Test (OFPT) was used (Oxford Language Centre), which contained 26 questions related to French grammar. This test was also completed in a pencil and paper format to avoid internet connection challenges. The full fifty-question test was used in a pilot study but was found to be taxing for participants of the target age group. The length of the OFPT, when completed with the background questionnaire, lead to a high level of frustration for some participants who then seemed to rush through questions at the end, potentially decreasing accuracy. The questions retained for the main study are available in Appendix B. No participants performed at ceiling, further justifying the use of a shortened version of the test. The questionnaire and grammar test were completed by participants individually, but participants were in the

same room as the person with whom they completed the production tasks. The researcher emphasized to participants that the results of the test were only for use within the study and would not be shared. Most participants completed the test without talking, but one pair began comparing answers and was asked not to do so. Comparison of the answers of this pair shows differences even on the answers discussed.

5.3 PARTICIPANTS

A total of 42 participants were recruited for the study by friend-of-a-friend method and ranged in age from 18 to 27 years old, including 20 males and 22 females. All participants resided in Meknes, Morocco at the time of the study. Three participants were in their final year of secondary education, while four participants had completed university studies and were employed or seeking employment. The remainder were university students at the time of recording. See Table 5.1 for a breakdown of age, gender, education level and field.

The high inclusion of English majors is an artifact of the recruitment method. These participants occasionally used English in their conversations, but most used less English than French. These students were sometimes eager to practice their English and therefore spoke to the researcher in English as well. Extra emphasis was made to this group to speak to their partner as usual. One pair of female English majors (P11251 and P11252) was excluded from analysis for near-categorical use of English in both tasks. One pair of male participants (P11211 and P11212) was similarly excluded from analysis for a high rate of Amazigh use which, while representative of their conversation style, lies outside the scope of this study. The use of other languages was a known possible outcome of the task design and is not included in the current analysis.

Table 5.1 Participant ethnographic data

Speaker	M/F	Age	Highest degree	Current Status	Current major/career field
P10281	F	26	3-year degree	Employed	Educational assistant
P11132	F	25	3-year degree	Undergraduate	Arabic Literature
P11131	F	24	3-year degree	Undergraduate	Geography
P12271	F	24	3-year degree	Undergraduate	Electromechanical Engineering
P10282	F	24	3-year degree	Employed	French/Arabic teacher
P11251	F	23	2-year degree	Undergraduate	English
P12272	F	23	3-year degree	Undergraduate	Materials Engineering
P12261	F	23	secondary	Undergraduate	Industrial Engineering
P11021	F	22	2-year degree	Unemployed	Business Agent
P11225	F	22	secondary	Undergraduate	English
P11293	F	22	2-year degree	Undergraduate	English
P12262	F	22	secondary	Undergraduate	Industrial Engineering
P11052	F	21	2-year degree	Undergraduate	Natural Sciences
P10232	F	21	secondary	Undergraduate	Humanities (not specified)
P11252	F	21	secondary	Undergraduate	English
P10301	F	20	secondary	Undergraduate	Law (Arabic)
P11226	F	20	secondary	Undergraduate	English
P11294	F	20	3-year degree	Undergraduate	Sociology
P10231	F	18	secondary	Undergraduate	Geology
P11051	F	18	secondary	Undergraduate	Economics
P11022	F	18	none	Secondary	Literature
P10302	F	18	secondary	Undergraduate	Law (Arabic)
P10261	M	27	3-year degree	Undergraduate	Bank Administrator
P11212	M	27	2-year degree	Undergraduate	English
P11211	M	25	2-year degree	Undergraduate	English
P11201	M	24	secondary	Undergraduate	English
P12231	M	23	secondary	Undergraduate	Mechanical Engineering
P12232	M	23	secondary	Undergraduate	Mechanical Engineering
P11224	M	22	2-year degree	Undergraduate	English
P11223	M	22	secondary	Undergraduate	English
P11214	M	21	secondary	Undergraduate	English
P11291	M	21	secondary	Undergraduate	English
P11271	M	21	secondary	Undergraduate	English
P11213	M	21	2-year degree	Undergraduate	English
P11272	M	21	secondary	Undergraduate	English
P11202	M	20	secondary	Undergraduate	English
P11222	M	19	secondary	Undergraduate	English
P11292	M	19	secondary	Undergraduate	English
P11221	M	18	secondary	Undergraduate	English
P10262	M	18	secondary	Undergraduate	Economics
P12031	M	18	none	Secondary	N/A
P12032	M	18	none	Secondary	N/A

Two sets of speakers failed to submit a written IM task: P10301, P10302, P12271, P12272. The spoken production of these speakers was used, but cannot be compared to their written language. Speakers P11051 and P11052 wrote completely in French as P11052 reports not being able to read Moroccan Arabic written in the Latin alphabet and for this reason their written task was excluded from the analysis. The IM chats of speakers P11213, P11214, P11221 and P11222 was similarly excluded due to high use of English; the first pair used 96% English while the second used 72% English. In total, the production by fourteen pairs of speakers (seven male, seven female) was analyzed for the written IM task.

Due to a technical malfunction, the recording of the spoken task of P12231 and P12232 contained only ten minutes and was not included due to the disruption to the conversation caused by the malfunction. This leaves recordings from eighteen pairs (8 male, 10 female) analyzed in the spoken conversation task.

5.4 DATA ANALYSIS

The two production tasks provide a corpus from which the current results are drawn. The written corpus consists of the full IM conversations of the included 14 speaker pairs as described above for approximately 13.5 hours of written MCA and 9026 words. The spoken corpus consists of one 15 minute conversation excerpt from each of the 18 speaker pairs, or 4.5 hours and 41,247 words. The excerpts of spoken conversation, chosen by the researcher, never included the first five minutes of the conversation in order to minimize any self-consciousness on the part of participants due to the presence of the digital recorder. In an effort to keep at least one conversation topic consistent, all of the conversation excerpts included discussion of education. All participants spontaneously

included this subject in their conversations, although the amount of time devoted to the topic varied by pair.

The spoken data was transcribed by three primary transcribers with assistance from native speakers of MCA where necessary. Fifteen of the spoken excerpts were transcribed by the researcher with an MCA native speaker assistant. For the transcription sessions, I prepared the .wav files for transcription in ELAN (Brugman and Russell 2004) and arranged meetings with a native speaker assistant. We then listened to the recordings together while the assistant clarified the pronunciation of any words or sounds that I did not understand in order to accurately represent them in the transcription system noted above. Any word or idiom meanings were also clarified by the assistant during this phase. This transcription arrangement was chosen for two reasons. First, no native MCA speaker could be found who was comfortable using transcription software. Second, the process allowed me to ask questions as they arose during the transcriptions and clarify any potential confusion immediately. However, this process was also extremely slow and required coordination with an assistant to schedule blocks of time two to five hours in length to work on transcriptions. The transcription process with an assistant took an average of 6 hours per fifteen minutes of conversation, or about 24 minutes of work per minute of transcribed conversation in addition to preparing and annotating the files.

One assistant worked with me in Meknes, Morocco but another was needed on my return to Austin as working together over the internet proved too complicated due to technical limitations, time zones, and her new job. A second assistant was found in Austin to work in a similar way, but returned home to Morocco before the transcriptions were completed. For this reason one 15-minute conversation excerpt was transcribed entirely by a native-like speaker of Moroccan Arabic who has spent over three years in-country. The final parts of 3 conversations were transcribed over a year after transcriptions began by a

native speaker of Moroccan Arabic who was completing his own linguistic research in the US and was able to independently transcribe the conversations. The full IM chats and the transcripts of the spoken conversation excerpts were then used to create a corpus for the current study, totaling approximately 50,000 words. Once the transcriptions were complete, I created a file for each speaker for each task, with many speakers having three files for the spoken task because the spoken files were transcribed in five minute chunks to make them easier to handle.

The Python programming language was used in the analysis due to its ease of use in natural language processing (NLP) applications. As a native speaker of English who speaks French at a near-native level and Moroccan Arabic at an advanced level, I created a Python program to compile a list of all of the words produced in each modality. I then created a copy of the list and deleted any words that were clearly not of French origin, and then went through a second time to verify that any that appeared to be of French origin actually were. I repeated this process for English starting with the original list. Any spoken lexical items with similar pronunciation in a variety of Arabic and French or English, such as radio or video as in (5.2), were assumed to be part of MCA.

5.2) ka-tbqa galəs, *connecté*, youtube, w kəda,
PRES-2nd-stay sitting, *connected*, youtube, and that

tšuf l-vidio
2nd-see the=video

"You stay sitting, connected, youtube, all that, you watch videos"
(P11291)

Lexical items that were orthographically ambiguous between French and English were all verified to be phonologically part of one language or the other. This was the case with words such as *culture*, /kʌltʃ͡ə˞/ in English or /kyltyʀ/ in French, and *pc* [computer], /pijsij/

in English or /pese/ in French. Context was used to distinguish the rare homonyms between Moroccan Arabic and a European language, such as *si*, meaning 'if' or 'yes' in French and 'mister' or part of a name in MCA. All words with apparent French or English origins were separately checked for inclusion in three bilingual Moroccan Arabic dictionaries (French: Mercier 1959, English: Harrell 1966, Spanish: Aguadé & Benyahia 2013) and one bilingual Standard Arabic-English dictionary (Wehr 1994) in addition to one in-depth study of borrowing in MCA (Heath 1989). If the word was absent from the Arabic presented by all five sources, it was considered a part of the suspected European language. Conversely, any European-origin words appearing in the Arabic section of a bilingual dictionary or the existing study of borrowings were considered established loanwords and thus fully Arabic.

Identification of the lexical items by language allowed detailed analysis of the French lexical items. Languages that lack a complete description, like the Arabic dialects, are impossible to analyze with many existing concordance software packages, and are difficult at best to analyze with many others due to the lack of a dictionary or syntactic tagging rules for these varieties. For this reason, I wrote a concordance-like program in Python to extract the French lexical items present in the corpus based on the list of French words created as described above. The Python program extracted the file name (always the speaker's numeric code), French lexical item, line number in which the lexical item appears in the original file, and full conversational turn in which the French lexical item was produced. A tab-separated values file was then created with this data in order to facilitate hand coding of constituent information using spreadsheet software. I wrote a Python program to count all occurrences of French lexical items in each file production and then compared the counts of French lexical items by person with the concordance output. In this

way I discovered an error in my program, which at first would skip any repetition of a single lexical item within a conversational turn. After this error was corrected, spot-checks were completed, including hand counts of French-origin lexical items, to ensure that the number of French-origin lexical items identified in the concordance matched the number in the original file. An extract of the concordance file as produced by the Python program is given in Table 5.2 below. If more than one French-origin lexical item was used in a single line of written chat or a single turn of spoken conversation, as in the first seven lines of the Table 5.2, the French lexical item and its associated conversational turn appear once for each lexical item.

Table 5.2 French-lexical items as extracted from spoken corpus with original context

File	File index	French lexical item	Conversational turn of lexical item and free translation
10231.txt	121	təypəy	lla ma ʕndiš ma ʕndiš TP physique ʕndi- ʕndi tlata TP-at
10231.txt	121	fizik	lla ma ʕndiš ma ʕndiš TP physique ʕndi- ʕndi tlata TP-at
10231.txt	121	təypəyat	lla ma ʕndiš ma ʕndiš TP physique ʕndi- ʕndi tlata TP-at
10261.txt	32	c'est	hiya c'est un client averti
10261.txt	32	un	hiya c'est un client averti
10261.txt	32	client	hiya c'est un client averti
10261.txt	32	averti	hiya c'est un client averti
11221.txt	6	pilote	bħal daba ila ža ši pilote ah-

This method of extraction was used in order to account for all French lexical items and to separately code the constituents found in French. I then identified the part(s) of speech of each French-origin lexical item and the type of constituent in which they appear by hand. An automatic part of speech parser was considered for this purpose, but was not used due to the low level of French present in the data and the inability of such programs to accurately analyze the French lexical items found in a primarily MCA context. The constituents were coded for constituent type, as seen in Table 5.3, whose contents correspond to that of Table 5.2.

Table 5.3 French constituents and type

File	File index	French constituent	Constituent type
10231.txt	121	TP physique	Bare noun phrase
10231.txt	121	TP-at	Bare noun
10261.txt	32	c'est un client averti	Complementizer Phrase
11221.txt	6	pilote	Bare noun

Additional morphosyntactic detail was added to the French-origin nominal structures found in the data, as described in Chapter 7.

In order to obtain a complete picture of the frequency of each lexical item, French words were also lemmatized. In this way, the French verb *accepter*, 'to accept,' was counted as occurring five times by three speakers, as in (5.3) below in their original written form.

5.3) a. ana [P10231] ***acptini*** [P10232]
I'm [P10231] *accept=me* [P10232]

"It's [name P10231], accept me, [name P10232] !!" (P10231)

b. flawal ana kant kan ***accépté*** bnadam *mais*
in=the=first me would.1st.SG *accept*.1st.SG people *but*

fhad tali walit ila ma3raftch dari
in=this following became if neg=know=neg.1st.SG boy

man ***acceptihch***
not=*accept*.1st.SG=him=NEG

"At first I would accept anyone but after that I changed so if I don't know a guy, I don't accept him." (P11226)

c. knt ghir kandir ***acc*** wahd nhar msht kolch li
would just do.1st.SG *accept* one day went all who

makanhdrch m3ahom bnat owlad *mais* had sa3a
NEG=speak.1st.SG=NEG with=them girls and=boys *but* this hour

dkhlo 3ndi wlad lblad bzaf obdit
entered space=my boys the=country many and=started.1st.SG

ghir	kandir	***acc***	hit	3arfahom
just	1st.SG.do	*accept*	because	knowing.F=them

"I would just accept [anyone] one day I got rid of everyone I don't talk to, girls and boys, but now a lot of guys from my area are on and I just accept because I know them." (P11225)

Speakers P10231 and P11225 use 'accepter' in a reduced form, although the meaning remains clear in the context – all of which in this case are referring to accepting people as friends on Facebook. Speaker P11226 uses a 'full' form of the verb, but the non-standard spelling of both instances, and apparent Arabic morphology on both instances of the verb, would make these unrecognizable in a typical automatic search for the terms. Frequency was then defined as the number of occurrences of the lemma in the corpus, tallied automatically through the software, regardless of spelling or morphology.

Finally, inclusion of the speaker information for each token of a lemma allows a simple measure of diffusion, defined here as [tokens of a word]:[unique users]. Thus the diffusion of *accepter* as seen in (5.3) above is expressed as 5:3. It is important to note that this ratio was never simplified in order to preserve a distinction between lemmas with greatly different frequency.

6. Results: Use of French by modality and social factors

This chapter examines the use of French lexical items by young adult speakers of Moroccan Colloquial Arabic in both written and spoken communication. The rate of French used and the syntactic constituents expressed in French are considered within each modality, as well as in relation to the social factors identified in Chapter 3 above: SEX, LANGUAGE ATTITUDE and FRENCH PROFICIENCY. These factors are explored in order to determine whether they impact the use of French among the target speakers. The inclusion of these factors provides an understanding of the use of CS in the traditional form, speech, as well as in the emerging literacy practices of the written form. In this way the current study examines the two primary settings in which the language is naturally produced. The comparability of this data between communication modalities is ensured by using the same speaker pairs for both tasks.

This analysis includes several novel aspects: the inclusion of both spoken and written CS data from the same individuals has not been previously undertaken in any language pair, and written data has never been examined for the current language pair as written MCA is a recent phenomenon.[18] However, as monolingual language use differs considerably in each modality, it is important to understand how CS practices may be similar or different in these modalities. The current data then aims to identify basic differences in the rate and structure of French used in each communication modality. Any such differences will be noted in this chapter in relation to the extralinguistic factors that may affect them.

[18] This topic merits a full investigation in its own right to describe the writing system of native MCA speakers and note any implications that may have on previously proposed analyses of the grammar of the dialects.

It is again emphasized that examination of rate of French use is an innovation in the study of Arabic-French CS; while past studies note the number of constituents in French (Bentahila & Davies 1995), the number of switches (Sayahi 2011a), or the number of the most common switches (Nait M'Barek & Sankoff 1988) no studies report these findings in light of the amount of Arabic employed by the same speakers. This simple detail makes it impossible to confidently compare the amount of French and CS used by speakers across studies. It is hoped that establishing a baseline of rate of French will allow comparison with other corpora in order to investigate questions such as change over time and differences between speaker communities. Any significant differences in rate of French between speaker groups based on extra-linguistic factors would also highlight the importance of this measure in an analysis of CS.

The consideration of constituents switched is common in studies of CS syntax and provides a way through which comparisons may be drawn between the current study and existing research on Arabic-French CS as well as other language pairs. As seen in Chapter 3, it has been assumed that the distinction between inter- and intra-sentential switching is indicative of speaker proficiency: speakers highly proficient in both languages are thought to utilize intra-sentential switches more often as they are competent in the individual grammar of each language. While this is true of the data between generations in Bentahila and Davies (1995), Sayahi's (2011a) analysis of speakers from a single generation does not reveal the same distinction. For this reason, greater emphasis is placed here on identifying differences in the rate and structure of CS in the current group in order to independently assess the linguistic factors associated with higher French proficiency in the current group.

Constituents switched may likewise illuminate the assumed differences in use of CS between male and female speakers. The same analysis is applied to LANGUAGE

ATTITUDE; as it has been put forward that this factor may impact judgment of CS, speakers may also use different structures if they approve of the phenomenon. Should differences be found based on extra-linguistic factors, it will be evidence that CS structure is not based purely on language-internal factors.

6.1 EFFECT OF COMMUNICATION MODALITY

One major contribution of this dissertation is the unprecedented comparison of written and spoken informal conversation between the same speaker pairs. This section gives an overview of the tasks as a whole to capture the basic differences and similarities in CS between written and spoken modalities. As noted in chapter 4 above, written language use is consistently found to be structurally and functionally different from spoken language, but how such differences may be realized in the CS of this language pair has not previously been examined. In fact, there are no extant studies on the spoken and written CS of the same users in any language pair to my knowledge. It is expected that differences will be found between the modalities, but the nature of these differences cannot be predicted. Written language is often more formal, but how this might be expressed when speakers use their dialect in Romanized Arabic is unknown due to the entirely non-standard nature of the written dialect and of the alphabet used to represent it.

One effect of communication modality is the corpus size for each task. The written data, taken from hour-long chat conversations by 14 speaker pairs, contains 9026 total words, or an average of 322 words per speaker. As most conversations are continuous conversations without gaps, indicated by the time stamps of each turn, this low number is indicative of the participants' typing speed and not of breaks in the conversation. By contrast, the spoken corpus is much larger; the 15 minutes transcribed from each of the 17

speaker pairs comprises 41,247 words, or an average of 1197 words per speaker, nearly four times as many words as the written conversations. The spoken data thus provides more language production for analysis despite the shorter time span over which the data for this task was collected.

This section compares and explores aspects of the two modalities in three subsections: first, the overall rate of French is considered in 6.1.1, followed by a description of the French constituents employed in 6.1.2. The section concludes with a discussion of the overall implications of the findings on modality in 6.1.3.

6.1.1 Rate of French lexical items

Rate of French, as described in chapter 3 above, is a useful measure for assessing participation in CS but it is rarely included in research. Analysis of the rate of each language provides a snapshot of overall use that can be useful in comparisons between data sets. In the corpus under investigation here, it is also a useful point of comparison for the use of CS between communication modalities. The first step of the analysis was to determine the proportion of the written data present in French. The lexical items from the full corpus were identified and tallied as described in Chapter 5. In the sections that follow, the rate of French lexical items used by speakers is examined, in turn, in informal Instant Messaging (§6.1.1.1) and spoken (§6.1.1.2) conversation. A brief discussion of rate and modality follow in §6.1.1.3.

6.1.1.1 Rate of French in Instant Messaging chat

The IM chat of the 28 participants included in the written corpus comprises 9026 words, of which 862 words, or 9.6%, are of French origin and absent from the bilingual dictionaries and study on borrowing described in §5.4. Many French lexical items are repeated, with 302 lemmas realized in the written data. Table 6.1 gives the overall use of

French used in the written corpus, including rate as the percentage of French words out of the total words and standard deviation of the rate among the speakers as a means to

Table 6.1 Average French token counts and French rate in IM chat

	Count	Speaker mean	Speaker median	Range	Standard deviation
Total words	9026	322.4	318.5	120-623	145.3
French tokens	862	30.8	22.5	3-128	28.4
French rate	--	9.6%	8.2%	1.2-30.1%	7.9%

summarize the differences in use found between individuals. As evidenced here, the vast majority of the IM conversations are in Arabic with a low rate of French use. The summary statistics indicate a vast degree of variation in the data, evidenced by the broad range and large standard deviations. The comparison between the median and the mean, as the latter is slightly lower than the former, indicates that the distribution of French rate is slightly skewed toward higher use of French. There are in fact a few high outliers as seen in the data by speaker, given in Table 6.2, and the general tendency toward Arabic is clear. This table shows the total words by each participant as well as the count of lexical items in French, given with the rate of French expressed as a percentage. The use of English is similarly included for comparison. The use of French in written conversation varies widely among participants, ranging from three words by P10281, shown in (6.1) to nearly a third of the conversation by P12261, seen in (6.2).

6.1) makykhlsna hta l 6 fchhar bach
NEG.PRES.3rd.us until the=sixth in.month so.that

mantlbouch *avance*
NEG.1st.ask.PL.NEG *ahead*
"He doesn't pay us until the 6th of the month so that we can't ask in *advance*"
(P10281, written)

6.2) *nn elle 'est pa connecté*
no she is not connected
"*No, she's not connected* [online]" (P12261, written)

In fact, P12261 is a student in Engineering, as are all four of the highest users of French in written IM. This field of study either attracts students who use more French or encourages them to do so; the current data does not indicate a reason for this trend.

CS with English in writing is therefore rare for all speakers; the use of English is much less common than French since 23 of the 28 speakers do not use any English at all. The five speakers who do employ English lexical items all do so at a lower rate than their use of French, with the highest rate of use at 5.9% compared to 10.1% of his writing in French. In fact, all participants who use English are English majors. Nevertheless, three English majors do not use any English, suggesting that factors other than area of studies impact use of English. French clearly maintains a privileged role in the speech of the participants as it is used by every speaker and represents a greater percentage of the data than English for all individuals who are included in this analysis[19].

[19] Two pairs of participants were excluded from the written analysis for near-exclusive use of English. These speakers report always using English with each other in this setting as a form of practice.

Table 6.2 Rate of French and English in IM chat by speaker

Dyad	Speaker	Total words	French words	% French	English words	% English
1	P10231	389	21	5.40%	--	--
1	P10232	317	20	6.31%	--	--
2	P10261	465	46	9.89%	--	--
2	P10262	623	51	8.19%	--	--
3	P10281	249	3	1.20%	--	--
3	P10282	352	4	1.14%	--	--
5	P11021	163	5	3.07%	--	--
5	P11022	137	16	11.68%	--	--
7	P11131	384	32	8.33%	--	--
7	P11132	200	7	3.50%	--	--
8	P11201	730	40	5.48%	--	--
8	P11202	120	10	8.33%	--	--
11	P11223	142	11	7.75%	--	--
11	P11224	320	22	6.88%	--	--
12	P11225	561	30	5.35%	--	--
12	P11226	373	32	8.58%	--	--
13	P11271	253	22	8.70%	--	--
13	P11272	246	38	15.45%	--	--
14	P11291	169	17	10.06%	10	5.92%
14	P11292	253	24	9.49%	4	1.58%
15	P11293	341	23	6.74%	5	1.47%
15	P11294	330	25	7.58%	4	1.21%
16	P12031	171	11	6.43%	--	--
16	P12032	169	9	5.33%	--	--
17	P12231	338	66	19.53%	--	--
17	P12232	386	76	19.69%	5	1.30%
18	P12261	415	128	30.08%	--	--
18	P12262	317	81	25.55%	--	--

6.1.1.2 Rate of French in spoken conversation

The data from spoken conversations, collected from 17 pairs of participants, contains a total of 41247 words, of which 2064 are of French origin and unattested in Moroccan Arabic dictionaries, totaling 542 unique lemmas. Table 6.3 gives the descriptive

statistics of French used in the spoken corpus, including rate as the percentage of French words out of the total words and standard deviation from the stated averages.

Table 6.3 Average French token counts and French rate in spoken conversation

	Count	Speaker mean	Speaker median	Range	Standard Deviation
Total words	41247	1213.1	1179	589-1935	412.1
French words	2064	60.7	32	4-396	76.7
French rate	--	5.2%	2.6%	0.5-34.0%	6.8%

The spoken use of CS varies even more than the written use, visible in the distance between the mean and median in addition to the relatively large standard deviation. From these observations, it is clear that most users employ very little French and half use less than 2.6%, as seen in (6.3).

6.3) l-ustad dyal- dħħkni ustad dyal la statistique
the.teacher of- laughs.me teacher of the statistic
"The teacher- the statistics teacher cracks me up" (P10231)

Combined with the large standard deviation, it is clear that a small number of speakers are outliers with a substantially higher rate of French use. The use of French is given by speaker in Table 6.4. and illustrates this fact; 25 of the 34 participants use less than the mean rate of French, while four participants are notable outliers, using more than twice the average rate of French lexical items. There was no sound reason to assume that the outliners belong to a different population than the rest of the participants and for this reason they were not excluded; in a larger sample, they would likely not be outliers. The lowest rate of French is used by P10281 at 0.50%, with two other speakers using French lexical items at a frequency of less than 1% in their speech, illustrated in (6.4).

6.4) hadi ɣadi tʕaṭik waħəd l'avantage aħsən
this.F going 3rd.F.give.you one the=advantage better
"This is going to give you an advantage, it's better." (P10281)

This use of a singe noun, with an ambiguous article, contrasts sharply with the highest rate of French in speaking, 33.99%, which was produced by P11052 who seems to alternate freely between languages as in (6.5).

6.5) wila kənti ka-tdiri des stages ula quelque chose
and.if were.2nd PRES=2nd.do.F some internships or some thing
"and what if you were doing some internships or something?" (P11052)

Her rate of French is also much higher than her conversational partner, who uses only 13.25% French. The presence of these extreme outliers means that statistical analyses must be applied only after transformation of the data to ensure that the data fits the assumptions of each statistical method. The spoken data was transformed using a logarithm function before all statistical analyses described below.

The number of participants using English is again low, but higher than in the written data with ten incorporating elements of English into their spoken conversations. As was the case with the written IM task, the users of English in the spoken modality are again English majors. In contrast to the written data, use of English in spoken CS exceeds that of French in the speech of P11271 and in the conversation between participants P11291 and P11292. The use of English is often in full clauses, unlike the majority of French lexical items, as noted in §6.1.2.2 The use of English merits investigation but lies outside the scope of the current study.

Table 6.4 Rate of French and English in spoken conversation by speaker

Dyad	Speaker	Total words	French words	% French	English words	% English
1	P10231	1626	41	2.52%	--	--
1	P10232	999	21	2.10%	--	--
2	P10261	791	120	15.17%	--	--
2	P10262	648	139	21.45%	--	--
3	P10281	804	4	0.50%	--	--
3	P10282	1355	27	1.99%	--	--
4	P10301	1935	33	1.71%	--	--
4	P10302	870	16	1.84%	--	--
5	P11021	1213	49	4.04%	--	--
5	P11022	977	24	2.46%	--	--
6	P11051	1404	186	13.25%	--	--
6	P11052	1165	396	33.99%	--	--
7	P11131	739	6	0.81%	--	--
7	P11132	1364	17	1.25%	--	--
8	P11201	1449	11	0.76%	2	0.1%
8	P11202	1339	18	1.34%	--	--
9	P11213	1504	38	2.53%	32	2.1%
9	P11214	1379	53	3.84%	48	3.5%
10	P11221	1902	132	6.94%	6	0.3%
10	P11222	914	78	8.53%	--	--
11	P11223	659	37	2.24%	--	--
11	P11224	1652	28	4.25%	--	--
12	P11225	1177	34	3.07%	16	1.4%
12	P11226	1109	43	3.65%	14	1.2%
13	P11271	826	140	7.54%	91	4.9%
13	P11272	1857	22	2.66%	53	6.4%
14	P11291	1214	31	2.55%	55	4.5%
14	P11292	1087	18	1.66%	30	2.8%
15	P11293	1115	22	1.97%	--	--
15	P11294	1181	31	2.62%	--	--
16	P12031	1467	24	1.64%	--	--
16	P12032	589	16	2.72%	--	--
18	P12261	2271	168	7.40%	--	--
18	P12262	666	41	6.16%	--	--

6.1.1.3 Comparison of rate in each modality

As a full group, participants use a higher rate of French lexical items in their writing, at approximately twice the rate of written French by mean, while the median remains more than three times higher in writing as seen in Table 6.5

Table 6.5 Mean, median and range of rate of French in writing and speech

	Written	Spoken
Mean % French	9.6%	5.2%
Median % French	8.2%	2.6%
Range of % French	1.20-31.1%	0.50-33.99%

However, four participants, marked with an asterisk in Table 6.6, go against this trend and use a higher rate of French in their spoken conversations. The lexical items used in each was further explored in order to identify a clear reason for the higher use of French in speech. There is no obvious reason for the comparatively high use of French by P10282 and P11021. However, the spoken data from speakers P10261 and P10262 includes a debate over economic terms, as P10262 is a university student in economics and P10261 works in banking, an except of which is seen in (6.6).

6.6) yʕny *les* *créances* ɣadi ykunu mʕ *les* *clients*
3rd.mean the debts going 3rd.m.be.pl with the clients

w *les* *dettes* ɣa-ykunu mʕ l-*fournisseur*
and the debts going=3rd.m.be.pl with the=supplier

"It means that the [lent] debts are going to be with the clients, and the [borrowed] debts are going to be with the supplier." (P10262)

This debate continued for some time due in part to P10262's insistence that he was right about the terms he learned, over his older brother's experience working at the bank. As indicated in the translation, the older brother was actually correct. For this reason, heavy use of economic terms persists for close to two and a half minutes of the 15 minutes transcribed for this pair and contains a large number of French nouns related to the topic.

Table 6.6 French token count and rate of French by speaker and modality

Dyad	Speaker	*	Written		Spoken	
			French tokens	% French	French tokens	% French
1	P10231		21	5.40%	41	2.52%
1	P10232		20	6.31%	21	2.10%
2	P10261	*	49	10.54%	120	15.17%
2	P10262	*	51	8.19%	139	21.45%
3	P10281		3	1.20%	4	0.50%
3	P10282	*	6	1.70%	27	1.99%
4	P10301		--[a]	--[a]	33	1.71%
4	P10302		--[a]	--[a]	16	1.84%
5	P11021	*	6	3.68%	49	4.04%
5	P11022		16	11.68%	24	2.46%
6	P11051		--[b]	--[b]	186	13.25%
6	P11052		--[b]	--[b]	396	33.99%
7	P11131		32	8.33%	6	0.81%
7	P11132		7	3.50%	17	1.25%
8	P11201		40	5.48%	11	0.76%
8	P11202		10	8.33%	18	1.34%
9	P11213		--[c]	--[c]	38	2.53%
9	P11214		--[c]	--[c]	53	3.84%
10	P11221		--[c]	--[c]	132	6.94%
10	P11222		--[c]	--[c]	78	8.53%
11	P11223		11	7.75%	37	2.24%
11	P11224		22	6.88%	28	4.25%
12	P11225		30	5.35%	34	3.07%
12	P11226		32	8.58%	43	3.65%
13	P11271		22	8.70%	140	7.54%
13	P11272		38	15.45%	22	2.66%
14	P11291		17	10.06%	31	2.55%
14	P11292		24	9.49%	18	1.66%
15	P11293		23	6.74%	22	1.97%
15	P11294		25	7.58%	31	2.62%
16	P12031		11	6.43%	24	1.64%
16	P12032		9	5.33%	16	2.72%

Table 6.6 (cont) French token count and rate of French by speaker and modality

17	P12231		66	19.53%	--[a]	--[a]
17	P12232		76	19.69%	--[a]	--[a]
18	P12261		128	30.08%	168	7.40%
18	P12262		81	25.55%	41	6.16%

a. Participant did not complete task
b. Participant's written chat was excluded from the analysis due to near-exclusive use of French
c. Participant's written chat was excluded from the analysis due to near-exclusive use of English

6.1.1.4 Discussion of Rate of French by modality

The data from these speakers clearly indicates a difference in rate of French use by modality. The participants use twice as many French lexical items in written chat than in spoken conversation on average. This could be a strong effect of the modality: as all speakers chose to write in Romanized Arabic, the use of what they call 'French letters' may trigger higher use of French lexicon. In addition to modality, the online nature of IM as a type of CMC may contribute to the use of French lexical items; a strong association between technology and the French language has been found in Morocco (Chakrani 2010, Ennaji 2002, 2005) and may encourage the use of French in interactions that are mediated by technology. However, the overall rate of French use remains very low in both modalities. In spoken CS, certain conversation topics and lexical items are often found to trigger CS (Muysken 2000, among others) and the same principle may be at work in the written modality.

An additional explanation may lie in the environment: Facebook's advertisement targeting tools show that 8.6 million users list Morocco as their country and 6.2 million of these, or 72%, are listed as the estimated target for advertisements in French, while just 3 million, or 35%, would be targeted in Arabic. Facebook does not reveal their methodology in arriving at this split, which indicates by the 107% total targeting that some users would be targeted using both languages. As many Moroccans anecdotally report using the

Facebook interface in French, this split by language is not surprising. The chat window in Facebook is embedded in the main webpage, which leaves visible other components of the website. If French words are visible while typing, it may subconsciously prompt speakers to include more in their own written conversations. Each of these factors may contribute to a higher use of French, with the overall result that a significantly higher rate of French is found in IM chat than in speech among current participants.

6.1.2 Constituents switched

Rate of French provides a clear measure of CS use that can be quickly compared with other language pairs and contexts, but does not indicate how individuals use each language or how the structures of two languages may come together. The research goals here are not only to describe the quantity of French used, but also the nature of its use, as previous studies have indicated that groups of speakers use French in distinct ways. This section gives an overview of the constituents switched in the aggregate for each task, followed by a comparison between tasks. A consideration of the constituents manifested as French insertions gives greater insight than does overall rates of switching into how speakers are using French within MCA.

6.1.2.1 Constituents switched in IM

The written corpus contains 599 French constituents. Examining the use of parts of speech in writing by speaker reveals that all speakers use nouns or noun phrases, while certain participants also incorporate a wide range of French lexical items from different grammatical categories into their writing. The parts of speech used in the written corpus are listed in descending order by the number of participants using each part of speech in Table 6.7.

Table 6.7 Written French constituents and number of speakers employing each

Constituent Type	French Constituents		Speakers (of 28)	
	Count	%	Count	%
Noun/Noun Phrase[20]	191	31.9%	28	100.0%
Oui/Non	122	20.4%	22	78.6%
Complementizer Phrase	66	11.0%	20	71.4%
Determiner Phrase	54	9.0%	22	78.6%
Verb	45	7.5%	18	64.3%
Conjunction	44	7.4%	17	60.7%
Adjective	28	4.7%	15	53.6%
Interjection	23	3.8%	15	53.6%
Adverb	17	2.8%	11	39.3%
Preposition/Preposition phrase	8	1.3%	6	21.4%
Number Phrase	1	0.2%	1	3.6%
TOTAL	**599**	100%		

The highest two categories are much more common than the others; Noun/Noun Phrase accounts for 32.2% of the data, such as *conversation* seen in (6.7).

6.7) mn b3d sa3a nssaliw had *conversation*
from after hour 1st.finish.PL this *conversation*
"In an hour when we finish this conversation." (P11294)

However, the use of nominal structures stands out numerically even more above the other constituent types if the Determiner Phrases, as in (6.8) is added to the Noun Phrase category, bringing the total to 41.2% of the data.

6.8) kifek m3a *les* *etudes*
How.you with *the.PL* *studies*
"How are your studies going?" (P11131)

The next highest category comprises the affirmative/negative items, Oui/Non, which represents 20.4%. Non-standard spellings are also included in this category and are quite common; of 47 negative items, only two appear with standard French orthography while the primary spelling is *nn* as in (6.9).

[20] Constituent categories under analysis are identified by the use of capital letters on each word.

6.9) nn mazaaal
"No, not yet" (P11022)

This is also noteworthy as *oui* and *non* have not previously been mentioned as items frequently switched within this language pairing. The Arabic forms of <yes/no> appear in the IM data, but at a lower rate, often written as 'ah' or 'la' as seen in (6.10) and (6.11).

6.10) a. ah ta ana kayjini ghir kdoub otfaliyat
yes even me pres.comes=me just lies and.childishness
"Yeah, me too, it seems to me just lies and childishness" (P11223)

b. iwa khasski t3elmi
yes need=you.F 2nd.learn.F
"Yes, you need to learn. (P11021)

6.11) la ghanhr9ha bhalak
no FUT.1st.burn=it.F like.you
"No, I'm going to burn it like you" (P10232)

The MCA forms of <yes> are /ah/ and /əywa/, as seen in (6.10a) and (6.10b) respectively. These are significantly different from the SA form /nʕm/, but have clear orthographic representations. The use of <no> may be less clear in writing as its phonetic form /lla/ is similar to that of the French definite article, /l̩a/ which has the standard orthographic form <la>. Geminate consonants are typically not represented in RMCA, rendering these two distinct forms orthographically indistiguishable. The similarity in form may lead speakers to write 'non;' however the different contexts in which the definite article and <no> are found would distinguish the two homographs and weakens this explanation.

Many of the constituents noted in the data are traditional syntactic constituents that do not require further explanation; however, certain trends within the data prompted the addition of other labels as well. The use of *oui* and *non* is treated separately from other tags or discourse markers in this analysis due to their common use in writing by

participants. This distinction is further justified by the difficulty in categorizing *oui* and *non* as they can also function as pro-sentences (Grover et al. 1975).

6.1.2.2 Constituents switched in spoken conversation

The spoken conversation data contains 1368 constituents, which were labeled as described in Chapter 5. Table 6.8 shows the parts of speech used in the spoken conversation corpus. Although not every speaker uses a noun or noun phrase, as they do in the written data, they all use a nominal structure, defined as a noun, noun phrase or determiner phrase. In fact, nominal structures constitute 75.7% of the data, dwarfing all other categories of constituents, The category Noun/Noun Phrase is always without a French determiner. However, the category Determiner Phrase may be used as a full DP in the utterance or, like in (6.12), may be a part of a larger MCA DP.

6.12) ayyəh ʕlaš? ħit had *les* *branches* hado aṣlan maši-
yes why? because DEM *the* *branches* these originally not-
"Yes, why? Because these *branches* here really aren't-" (P11222)

The example in (6.12) is headed by the MCA demonstrative determiner *had* and is followed by the second demonstrative *hado* to add emphasis. The fact that these structures differ in syntactically important ways leads to their separation in their primary analysis, yet their similarity as different ways of incorporating a French noun in MCA-French CS, justifies also viewing these constituent types as a group.

The extremely high use of nominal structures leaves little room for other constituents, which are spread throughout. However, the overall rate of use of a given constituent in the full corpus must be considered with the number of speakers who use the part of speech to find a more nuanced understanding of the use of the constituent types present in the data. For example, Complementizer Phrases (CP),[21] represent about the same

[21] This category is distinct from that of Complementizers, which includes only bare Complementizers

Table 6.8 Spoken French constituents and number of speakers employing each

	French Constituents		Speakers	
Constituent Type	Count	%	Count	%
Noun/Noun Phrase	628	45.9%	33	97.1%
Determiner Phrase	407	29.8%	32	94.1%
Verb/Verb Phrase	62	4.5%	22	64.7%
Adjective	58	4.2%	19	55.9%
Complementizer Phrases	50	3.7%	9	26.5%
Number/Number Phrase	42	3.1%	13	38.2%
Adverb/Adverb Phrase	39	2.9%	14	41.2%
Oui/Non	31	2.3%	13	38.2%
Preposition/Preposition Phrase	17	1.2%	7	20.6%
Conjunction	17	1.2%	12	35.3%
Interjection	7	0.5%	7	20.6%
Definite Article	6	0.4%	4	11.8%
Quantifier Phrase	2	0.1%	1	2.9%
Object Pronoun	1	0.1%	1	2.9%
Complementizer	1	0.1%	1	2.9%
TOTAL	**1368**	**100.00%**		

amount of overall data as Numbers/Number Phrases and Adverb/Adverb Phrases. However, while CPs constitute a similar proportion of the data, they are used by a smaller portion of the speakers with only 9 individuals employing them. In fact 35 CPs, or 70% of those found in the spoken corpus, are produced by a single pair of speakers (P11051, P11052) while the other 15 are employed by 7 other speakers. The count of speakers using the part of speech, then, indicate which constituents are used often by a small number of speakers rather than used infrequently by a large cross section of participants.

Although verbs and verb phrases make up less than five percent of the total French lexicon used in the spoken data, a variety of speakers employ them, as 22 participants use at least one French verb. It is noteworthy that only 11 of the 53 verbs appear as uninflected or with a French noun, as in (6.13); the vast majority (79%) show fully Arabic morphology seen in (6.14).

6.13) w bʕda tu sais bʕda organiser bʕda le temps dyalk
and after you know after to.organize after the time of=you
"and, like, you know, like, how to organize, like, your time" (P11052)

6.14) sfṭ lo invitation w huwa accepté
sent.1st to=him invitation and he accepted=it
"I sent him an invitation and he accepted it" (P11213)

When morphologically integrated into MCA, French verbs are typically treated as a native verb with a weak quadriliteral root, those whose final consonant is a semi-vowel (Harrell 1966, Caubet 1993, Ziamari 2008), seen in Table 6.9 with the French verb /klike/ *cliquer*, 'to click,' and MCA *ṣeqṣi*, 'to ask'. /klike/ does not occur in the current corpus, but its absence from the bilingual dicitonaries and Heath would lead it to be considered an instance of CS by the current standards. The verb integration pattern presented in Table 6.9

Table 6.9 French verb integration and weak quadrilateral verb pattern in MCA

	Perfective	Imperfective
First person sg	kliki-t (ṣeqṣi-t)	n-kliki (n-ṣeqṣi)
Second person sg	kliki-ti (ṣeqṣi-ti)	t-kliki (t-ṣeqṣi)
Third person masc. sg	klika (ṣeqṣa)	y-kliki (y-ṣeqṣi)
Third person fem. Sg	klika-t (ṣeqṣa-t)	t-kliki (t-ṣeqṣi)
First person pl	kliki-na (ṣeqṣi-na)	n-klikiu (n-ṣeqṣiu)
Second person pl	kliki-tu (ṣeqṣi-tu)	t-klikiu (t-ṣeqṣiu)
Third person (masc/fem) pl	klika-w (ṣeqṣa-w)	y-klikiu (y-ṣeqṣiu)

Adapted from Tables 1, 2 in Ziamari (2008:113)

is so common that speakers may spontaneously insert new verbs into this frame (Heath 1989:38). Spontaneous use of this verb pattern is evidenced in the speech of P11201, who twice surprises his friend with novel uses of an English verb with MCA morphology, given in (6.15).

6.15) P11201: t-*foksi*
2nd.prs-focus
"(You) focus"

P11202: t-*foksi*? a saħəbi, trakkəz
2nd.prs-focus-fem? VOC friend=my 2nd.prs-focus
"Focus English? My friend, focus Arabic."

Here it is clear that P11202 does not approve of his friend's adaptation of the English *focus* into MCA, although the accuracy of his suggested substitution indicates that he understood it effortlessly. For these reasons, all instances of foreign verbs with MCA morphology were noted unless the verb was found in the sources listed in Chapter 5. The low overall use of verbs in the corpus created for this study prevents further analysis of whether morphological integration is an indicator of borrowing into MCA; however, such an analysis is necessary if a sufficiently large corpus can be found.

6.1.2.3 Comparison of constituents switched in each modality

The types of constituents used in each modality are similar, but more variety is found in the Spoken conversations. In order to compare directly between modalities, the constituents that appear in each modality are represented as a percent of all the French constituents in that modality in Table 6.10.

Three constituent types show large differences between modalities. Determiner Phrases display the greatest difference between modalities as they constitute nearly a third of all constituents in speech, but are only 9% of the constituents in the written corpus. This is surprising as the phonetic similarity of the Arabic and French definite articles (i.e. /l/, /lə/, see the discussion of noun morphosyntax in §7.3 for further detail) would be expected to lead speakers to pronounce French words with an Arabic article, whereas the orthographic differences between the article in each language would likely be more salient in the written modality, leading participants to use the French article more often.

Table 6.10 French constituent types as percentage of all French constituents in written and spoken tasks

French constituent	Written	Spoken
	% PoS	% PoS
Noun/Noun Phrase	31.8%	45.9%
Determiner Phrase	9.0%	29.8%
Oui/Non	20.4%	2.3%
Complementizer Phrases	11.0%	3.7%
Verb/Verb Phrase	7.5%	4.5%
Conjunction	7.4%	1.2%
Adjective	4.7%	4.2%
Interjection	3.8%	0.5%
Number/Number Phrase	0.2%	3.1%
Adverb/AdverbPhrase	2.8%	2.9%
Preposition/Preposition Phrase	1.3%	1.2%
Other	--	0.7%

Conversely, the proportion of <yes> and <no> as French constituents in the written data is nearly ten times higher than the spoken, indicating some type of association between the French forms and items in written online communication, which may explain why the use of these forms in French was not noted in past studies. The third notable difference between written and oral modalities is between Nouns and Noun Phrases; the latter are more common in spoken conversation where they represent 13.3% more of the French lexical items used. The total rate of combined nominal structures (DP, bare nouns, and bare NPs) in the spoken data is 75.7% whereas they comprise only 40.8% of the written data. The greater use of French constituents seen in the written data is then not simply an overall increase as the proportion of constituent types changes greatly between the two modalities.

The lower proportion of nominal structures in the written data means that other constituent types make up the difference. Two constituent types display notable use in the written corpus, but are almost absent from the spoken: conjunctions and interjections. These constituents can be described as extra-sentential elements, which are often believed

to be easier to incorporate in CS as these elements can be used at the edges of utterances, without need to access the grammar of either language to incorporate them syntactically. The higher use of extra-sentential elements in written data may be a way in which speakers with a lower proficiency in French or who do not normally participate in CS can make use of French lexical items, as would be predicted by Poplack (1980). This is examined below in §6.3. Another possibility is that both of these constituent types are indicative of the greater opportunity for planning presented by After determiner phrases and nouns/noun phrases, only numbers and number phrases are notably more common in the spoken corpus than in the written. However, this may again be an artifact of modality as many of the numbers in the written chat were expressed as digits and therefore cannot be classified for language in the written modality.

6.1.3 Discussion of effect of modality

It is clear that modality of communication has an effect on use of French, as both rate of the use of French and a count of the constituents used in French differ greatly between the modalities. The rate of French items in the written data (9.6%) is nearly twice that of the spoken data (5.2%), indicating an important difference in modality for this language pair and age group. Almost all speakers decrease their use in speech, a trend that is rarely contradicted, and these exceptions may be the result of conversational subjects that are highly associated with French. The substantial differences between modalities justifies a separate analysis of each. For this reason, the results from the written and spoken data are reported separately in the following analysis of the social factors under consideration.

The high use of nouns, noun phrases and determiner phrases in each modality is not surprising, but the difference between the rate of these categories and past findings on

spoken data is striking. The approximate rate of nominal structures in four prior studies on spoken Arabic-French CS is found in Table 3.1 above and repeated for convenience with the addition of the current data as Table 6.11. Caution must be taken in this comparison as each author describes use of French differently, and thereby may categorize certain cases differently. However, descriptions are similar enough to validate comparisons between studies. The ratio of use of nominal structures in French in the current spoken sample is most similar to Sayahi's results for male speakers without a university education, who used 71.3% nominal structures. As noted in §3.3, these results raise the question of whether a higher rate of nominal structures may be indicative of lower proficiency in French, a possibility that is explored within the current sample in §6.3 below.

Table 6.11 French nominal structures as a percentage of all French constituents in five Arabic-French CS corpora

Study	**Participants**	**% NS**
Nait M'Barek & Sankoff 1988[a]	Moroccans residing in Montreal	35.1%
Bentahila & Davies 1995 [b]	Older generation	22.2%
Bentahila & Davies 1995 [b]	Younger generation	62.8%
Redouane (2005)	Moroccans residing in Quebec	28.1%
Sayahi 2011a[c]	University-educated Tunisian males	59.0%
Sayahi 2011a[c]	University-educated Tunisian males	63.8%
Sayahi 2011a[c]	University-educated Tunisian females	71.3%
Current	Spoken data	75.7%
a – data from 'Ns followed by Arabic,' 'Ns at end of phrase' (1988:148) b – data from 'for a whole NP', 'between Det and Det', 'between Det and N' (1995:81) c – data from 'Bare Nouns,' 'NPs' (2011a:126)		

Another potentially important factor that arose in the analysis is area of study. The four engineering students each use a much higher rate of French compared to the average participant. These speakers seem to have a substantially different experience with French. In statistical terms, it appears that two subpopulations are included in the same sample. A sound analysis must include a single subpopulation, or representative samples from across

the full population. As more data cannot be gathered at this time, the engineering students are considered a separate population from the rest of the participants and are not included in the following analyses of extra-linguistic factors. The differences between students in engineering and other fields of study is an important issue, but cannot be adequately addressed by comparing the current sample of four students in engineering with the 32 participants from other fields of study.

While both modalities exhibit a high rate of nouns in French, more differences can be observed than similarities. The written data contains fewer constituent types and is more balanced among these types as the variance between the constituent types is smaller than in the spoken data. Noun/Noun Phrases is the largest constituent type in both modalities, but nominal structures combined account for 34.9% more of the data in the spoken corpus. French lexical items in the written corpus are more evenly divided between parts of speech: six different parts of speech contain at least five percent of the written French, while only two parts of speech reach that level in the spoken data. The extreme difference in use of 'oui' and 'non' is the most obvious constituent affected by modality. The fact that there are fewer numbers used in French in the written data is also a direct effect of modality: in the written corpus, numbers are often expressed as digits and are therefore without language. Verbs and their morphology are of particular interest as Ziamari (2008) noted that French verbs may be directly inserted into MCA with MCA morphology through the common borrowing routine noted in §6.1. Most of the verbs in the written corpus use MCA morphology with only one verb *je pense* ('I think') used in a fully inflected French form, and it is used just twice, once each by speakers of either sex.

6.2 EFFECT OF SEX ON FRENCH USE IN CS

In Chapter 3 it is noted that SEX as a sociolinguistic variable is commonly found relevant in sociolinguistics studies of language use. In the Arab world, many researchers have suggested that women use more CS (Ibrahim 1989, Lawson and Sachdev 2000, Daoud 2011) although there is little data with which one can support or refute such a claim. Sayahi (2011a), whose results indicate that women use CS at about the same rate as men, is the only prior study in Arabic-French to explore the observed use of French in a quantitative way. Chakrani's findings on reported use also suggest similarity in CS between the sexes. However, these studies also provide evidence that the way in which speakers of each sex use CS may differ (Sayahi 2011a) as well as the pronunciation (Ziamari 2008). It is possible that they way in which women use CS is more salient, thus reinforcing the stereotype that they employ a greater amount of CS. In this section I explore the written and spoken corpus to identify whether SEX is observed to play a role in use of French in the current data.

6.2.1 SEX and rate of French

As in many sociolinguistic studies, it is likely that notable differences will be found between the sexes in rate of French. While not a true rate, Sayahi's (2011a) report of the number of French items used by each speaker in fifteen minutes is the only study that compares amount of French used in Arabic-French CS. Sayahi (2011a) found that university-educated males and females used a similar number of French lexical items in conversation by considering the raw count of French switches. While this measure gives an overall idea of amount of French used, it does not indicate the proportion of French used by interviewees as no total word count is given. For this reason, a percentage is given here in addition to a count in order to take into account other potential differences, such as rate of speech.

6.2.1.1 SEX and rate of French in Instant Messaging

The use of French in IM is examined first, as above. No expectations specific to this modality and extra-linguistic factor exist due to the lack of previous studies of written Moroccan Arabic. Table 6.12 shows the average total word count, the average count of French words, and the average French percent by sex. The overall rate of French for the data included in the analysis of social factors is provided for reference as the data presented in these sections differs from that given in §6.1 due to the exclusion of a pair of participants who are engineering students, as noted above.

Table 6.12: Average rate of French lexical items in IM by sex

Sex	Average words all languages	Average French-origin words	% French-origin
Female	316.3	18.5	5.9%
Male	305.1	26.1	8.9%
Overall	310.7	22.3	7.4%

The overall rate of written production is very close between the sexes, with total average word counts just eleven words apart. The raw use of French differs with males using an average of 7.6 words more than their female counterparts. When viewed in light of the limited written production by all participants, these extra words in French result in a 3% greater use of French by male participants in writing compared to their female counterparts.

I carried out a linear mixed effects analysis of the relationship between rate of French and SEX in R (R Core Team 2012) using lme4 (Bates, Maechler & Bolker 2012). The fixed effect in the model was SEX, with speaker pair as a random effect. Rate of French in writing was the dependent variable. Visual inspection of the data did not indicate any violations of normality. A likelihood ratio test of the model including SEX and the model

without this factor was used to obtain a P-value. This result indicates that speaker sex significantly impacts the use of French in the written data (p= 0.0156).

6.2.1.2 SEX and rate of French in spoken conversation

As all previous studies on CS between French and Arabic are on spoken language, the analysis of the current data begins with the spoken conversation data. The overall comparison between sexes in speech is seen in Table 6.13. The averages for the subset of the data included in the analysis of social factors is given alongside the average by sex for reference, as the data presented in these sections differs from that given in §6.1 due to the exclusion of a pair of participants who are engineering students, as noted above.

Table 6.13 Average rate of French lexical items in speech by sex

Sex	Average words all languages	Average French-origin words	Average French rate
F	1189.6	59.4	4.9%
M	1204.8	56.6	5.4%
Overall	1197.2	58.0	5.1%

In spoken language, the rate of French use by female participants is only 0.5% lower than that of male participants. Males and females are equally likely to use French lexical items in their speech, despite the fact that males use more in written communication.

I carried out a linear mixed effects analysis of the relationship between rate of French and SEX was in R (R Core Team 2012) using lme4 (Bates, Maechler & Bolker 2012). The fixed effect in the model was again SEX, with speaker pair as a random effect. Rate of French in speaking was the dependent variable. Visual inspection of the data indicated a violation of normality, leading to a data transformation by a logarithm function in order to meet the assumption of normality of this test. A likelihood ratio test of the model

including SEX and the model without this factor was used to obtain a P-value. This result indicates that speaker sex does not impact the use of French in the spoken data (p= 0.47).

6.2.1.3 Discussion

The observed use of French lexical items by SEX diverges notably from the popular stereotypes, which predict higher use of French by female speakers. The overall findings with regard to modality indicate that both female and male speakers of MCA use a lower rate of French lexical items in speech than in writing, as summarized in Table 6.14.

Table 6.14: Average rate of French lexical items by sex and modality

Sex	% French-origin words in writing	% French-origin words in speech
Female	5.9%	4.9%
Male	8.9%	5.4%
Overall	7.4%	5.1%

While previous results are confirmed regarding the similarity of the rate of CS between sexes in speech (Sayahi 2011a, Chakrani 2010), it is revealed that male speakers use a significantly higher rate of French lexical items in writing. These findings contradict the traditional assumption that females use more CS (Ibrahim 1988, Daoud 2011, Lawson & Sachdev 2000). The reasons for the high use of French observed in the written modality are not readily apparent; it may be that the use of the Roman alphabet subconsciously prompts users to use more French, or perhaps the act of looking at a computer screen and using a keyboard, which may or may not contain the Arabic alphabet, leads users to do so. The use of technology to communicate may subtly encourage speakers to discuss subjects that are more associated with French, as the domain of technology itself is often found to be associated with French (Ennaji 2002, Chakrani 2010). This latter possibility may also be related to the higher use of French lexical items by males in writing; reading the

conversations leaves the impression that they discussed a greater number of subjects traditionally related to French, primarily technology, such as in (6.16).

6.16) wakha tadir liha al misajour taykhadmo
Ok 2nd.do to=it the update pres.3rd.work.pl

les jeux mn disque dur
the games from hard disk
"Ok, do an update on it and the games work from the hard drive" (P11221)

A thorough analysis of topic and the potential use of more technological terms by male participants would require identification of the semantic domains discussed in the conversations and is outside the scope of the current study. However, future analysis of this question is clearly warranted.

6.2.2 SEX and constituents switched

The results regarding rate of French settle the question of whether one sex uses more French than the other, but it may be that gender-based stereotypes exist due to the ways in which individuals use CS. If a certain constituent type is more aurally salient to the listener, stereotypes may develop based on the use of these constituents and not overall incorporation of French-origin lexical items. This possibility is suggested by the differences in the constituents used by sex as noted by Sayahi (2011a) and is explored here for this reason.

6.2.2.1 SEX and constituents in Instant Messaging

Table 6.15 shows the constituents used in IM chat by participants, divided by sex. The tokens counts of each constituent type are listed as well as the percent of the given constituent type out of all French constituents produced by participants of that sex. The overall information is also included because the data included in this table differs from that in §6.1.2.1 above as the engineering students are excluded from the extralinguistic analysis.

Table 6.15 French constituent types as a percentage of switches by sex in writing

Constituent type	Females		Males		Both sexes	
	%	Token count	%	Token count	%	Token count
Noun/Noun Phrase	21.2%	39	43.1%	103	33.6%	142
Oui/Non	29.3%	54	15.9%	38	21.7%	92
Conjunction	12.0%	22	6.7%	16	9.0%	38
Complementizer Phrase	10.9%	20	5.9%	14	8.0%	34
Determiner Phrase	8.2%	15	7.9%	19	8.0%	34
Verb	6.0%	11	6.7%	16	6.4%	27
Adjective	2.2%	4	6.7%	16	4.7%	20
Interjection	6.0%	11	3.8%	9	4.7%	20
Adverb	3.3%	6	2.1%	5	2.6%	11
Preposition/Prepositional Phrase	1.1%	2	0.8%	2	0.9%	4
Number Phrase	--		0.4%	1	0.2%	1
Total Constituents		184		239		423

It was established in §6.2.1.2 that male participants use a significantly higher rate of French-origin lexical items in their written communication. The constituents involved also differ as 43.1% of constituents used by male speakers are Noun/Noun Phrase, twice the proportion of the same constituent type in the female data. The only other constituent type with a higher proportion of male use than female use is Adjectives, although this category is of relatively low frequency among all speakers (6.7% of the male constituents and 2.2% of the female). It is the higher use of the Noun/Noun Phrase constituents among males that accounts for the entire difference between French-origin constituents between male and female speakers. As Noun/Noun Phrases are content words and generally indicate the topic of conversation, this finding strengthens the suggestion above that the higher male use of French is due to the subjects discussed in this modality.

Female speakers employ a wider variety of constituent types as they use more Oui/Non, Complementizer Phrases, and Interjections. The use of Oui/Non by females may

appear in a turn with other words, seen in (6.17) but 19 instances of <oui> are the sole word in a turn, as are 8 uses of <non>.

6.17) oui bezaf
yes lots (P11021)

The Oui/Non constituent type is more common among female participants, but it is far from absent in the male data as it is the second most common constituent type for that group and may also be the sole lexical item in a written turn (9 instances of 'oui' and 3 of 'non'). However, the token count reveals that the differences between the use of Interjections and Complementizer Phrases is similar between the sexes in raw frequency, but the high use of Noun/Noun Phrase by male speakers reduces the proportion of all other constituents in their writing.

Despite these noticeable differences, several constituent types are used in similar proportions by male and female participants. Determiner Phrases, the third most common constituent type among males and fourth most common among females, represent a similar proportion of the data of each sex. Less common constituents are also distributed similarly between the sexes: Verbs, Adverbs, Prepositions, Complementizer Phrases, and Number Phrases.

6.2.2.2 SEX and constituents switched in spoken conversation

While no significant difference was found between the rate of French used between sexes in speech, it may be that the types of French constituents that speakers incorporate varies by sex. Table 6.16 shows the number of times a given part of speech is employed by sex. The percentages given represent the percent of constituents produced by members of each sex in a given part of speech with the total of all speakers included in the analysis of extra-linguistic factors in order to give a full picture of the data without the engineering students.

Table 6.16 French constituent types as a percentage of switches by sex in speaking

Constituents	**Female**		**Male**		**All speakers**	
	Token count	% of F constit.	Token count	% of M constit.	Token count	% of constit.
Noun/Noun Phrase	224	41.4%	335	49.9%	559	46.1%
Determiner Phrase	163	30.1%	203	30.2%	366	30.2%
Verb/Verb Phrase	13	2.4%	41	6.1%	54	4.5%
Adjective	15	2.8%	38	5.7%	53	4.4%
Complementizer Phrase	37	6.8%	8	1.2%	45	3.7%
Number/Number Phrase	29	5.4%	7	1.0%	36	3.0%
Oui/Non	19	3.5%	10	1.5%	29	2.4%
Adverb/ AdverbPhrase	12	2.2%	14	2.1%	26	2.1%
Conjunction	11	2.0%	5	0.7%	16	1.3%
Preposition/ Preposition Phrase	10	1.8%	5	0.7%	15	1.2%
Interjection	5	0.9%	1	0.1%	6	0.5%
Definite Article	--	0.0%	5	0.7%	5	0.4%
Quantifier Phrase	2	0.4%	--	0.0%	2	0.2%
Complementizer	1	0.2%	--	0.0%	1	0.1%
Grand Total	541	100.0%	673	100.0%	1213	100.00%

As the French constituents used in writing vary with sex of the speaker, it is not surprising that differences emerge by sex in spoken language as well. However, like the rates of rate of French in speaking are very similar, fewer differences emerge in the constituents used by sex in this modality. One rare constituent type used by males, definite article, is notable only for its absence from the female production data. A small number of lone Determiners, all of which are false starts or precede incomplete words, are produced by male participants as seen in (6.18).

6.18) hadok lli galt lk lli yallah ka-nqraw *les* *en-* a-
those that told.I to=you that now PRES.1st.study.PL the.PL en-
Those ones that I told you that just now we study *the* *en-* a-
(P10262)

Male speakers use more Noun/Noun Phrases at more than half of the constituent types they produced. Male speakers used 79 more Nouns/Noun Phrases than females, which is greater than the overall difference between all constituents used by the sexes. The use of Determiner Phrases between the sexes is similar, but male speakers use twice the rate of Adjectives, and close to twice the rate of Verb/Verb Phrases compared to the female participants. These four most common constituent types constitute almost all of the French constituents used by males as the remaining 8.4% of data is spread amongst 8 different constituent types.

Female speakers show the highest preference for nominal structures, which represent 71.5% of the French constituents used by this group. Also visible in Table 6.16 is a higher use of Full Clauses, Number/Number Phrases and Oui/Non than the male participants. However, these categories must be interpreted with caution; 35 of the full clauses by females in the data were produced by a single pair of speakers, P11051 and P11052. Numbers/Number phrases are similarly skewed by this pair as they are the source of 19 of the 29 female Number/Number Phrases in the data, seen in (6.19)

6.19) oui ta huwa kan bʕid ʕliya kənt kan naxŭd deux taxis
yes even he was far from.me was.I PAST 1st.take two taxis
"yes, it was also far from me, I used to take two taxis" (P11052)

This pair also produced 15 of the 19 uses of Oui/Non by female speakers. The presence of this pair alters our understanding of the data yet, unlike the engineering students, there is no clear justification for excluding these speakers as they are both university students in the same public universities as the other participants. In their written chat task, these two discuss plans to study in other countries, which may be a strong source of motivation to learn other languages, but as this issue was not addressed in the personal background survey, it is impossible to know if other participants share this motivation or not. For this

reason pair P11051/P11052 was retained in the data set. The strength of using a mixed model statistical analysis allows for including random variables; using pair as a random variable minimizes any effect created by this pair that is not echoed by other female participants. As all of the other constituent types account for about 2% or less of the total French constituents, there is no constituent type that is dominated by female participants; instead, their use of French is marked by the variety of constituents used.

6.2.2.3 Discussion of SEX and constituents

As the primary trends for each sex by modality have been acknowledged, the final consideration of SEX focuses on whether the differences between sexes in each modality are mirrored in the other modality, or if modality of communication impacts the use of French by each sex in a different way. Table 6.17 gives the proportions of constituents used in French by modality and sex to investigate this question.

Table 6.17 French constituent types as a percentage of switches by modality and sex

Constituents	**Written**		**Spoken**	
	Females	Males	Females	Males
Noun/Noun Phrase	21.2%	43.1%	41.4%	49.9%
Determiner Phrase	8.2%	7.9%	30.1%	30.2%
Adjective	2.2%	6.3%	2.8%	5.7%
Complementizer Phrase	10.9%	5.9%	6.8%	1.2%
Verb/Verb Phrase	6.9%	6.7%	2.4%	6.1%
Number/Number Phrase	0.0%	0.4%	5.4%	1.0%
Oui/Non	29.3%	15.9%	3.5%	1.5%
Adverb/Adverb Phrase	3.3%	2.1%	2.2%	2.1%
Conjunction	12.0%	6.7%	2.0%	0.7%
Preposition/Preposition P	1.1%	0.8%	1.8%	0.7%
Interjection	6.0%	3.8%	0.9%	0.1%
Definite Article	--	--	0.0%	0.7%
Quantifier Phrase	--	--	0.4%	0.0%
Complementizer	--	--	0.2%	0.0%
Grand Total	100.0%	100.0%	100.0%	100.0%

When considering the sexes in concert with communication modality, certain observations remain constant; male participants use higher rates of Nouns/Noun Phrases and Adjectives in their French than the female participants in both modalities. The proportion of N/NP rises notably for both sexes in speech, but the use of Adjectives is quite similar across modalities. While Determiner Phrases display a large difference between modalities, the proportion of Determiner Phrases for each sex is similar in a single modality. It is not clear why use of N/NPs should behave so differently from DPs, indicating that further analysis of the type of nominal structure used by the participants is warranted. This question is taken up below in Chapter 7.

Modality has an effect for both sexes in the use of certain constituents, confirming the overall results in 6.1 above. A consistent finding across modalities is that females use about twice the rate of Oui/Non as male participants, regardless of modality, and that all speakers use them much more in writing. Despite the higher rate of use of Oui/Non by females across modalities, the use of this constituent type drops dramatically for all speakers in the spoken corpus. Complementizer Phrases are also used more by females, and much more often in writing; this may be related to the uses of fixed greeting phrases such as *ça va?*, "how are you?" and *et toi?* or "and [how are] you?" which appear 20 times in the IM corpus with 15 of these uses by female participants.[22] The spoken data, 15 minutes drawn from hour-long conversations that never includes the beginning of a conversation, would then not be expected to contain these types of phrases. However, it is expected that inclusion of spoken greetings would not substantially increase the use of French in the spoken data.

[22] The phrase "et toi?" was always used in greetings to mean "and how are you?" Fore this reason it was considered to be a CP that has undergone ellipsis.

Sayahi (2011a) found that male and female speakers with a university education used a similar proportion of Determiner Phrases, both near 24%, but that female speakers used a higher proportion of Nouns at 40% of their French constituents switched. The current results show a higher rate of Determiner Phrases at 30% of the French constituents by sex, while the rate of Noun/Noun Phrases is similar to Sayahi's at 41.4%. However, in the current corpus this is lower than the male rate of Noun/Noun Phrases of 51.4%. Not only does the proportion of constituents switched vary by sex, but the trends diverge between the two data sets.

6.2.3 Discussion of SEX and French use in CS

The results contradict existing stereotypes regarding use of French by speakers of Moroccan Arabic. Females are observed to use the same amount of French as males in speech and in IM they use significantly less. The suggestion that males may use more French in written CMC due to the subjects they discuss, put forth in §6.2.1.2, is supported by the higher proportion of nouns, the content words that relate to conversation topic, that they use relative to female participants. The use of other French constituents by females may draw more attention to the lexical items of French origin despite the fact that they are used less often. Nouns are typically considered the easiest part of speech to borrow, along with tags and interjections. While the males incorporate French nouns into their speech, it may be that the use of noun insertions is less salient to listeners and the perception that they are using less French persists. By contrast, the female speakers' use of interjections and complementizer phrases may draw attention to their use of French. A related possibility is pronunciation; if male speakers tend to pronounce French-origin nouns with a strong Arabic accent the use of nouns may be less obvious. If female speakers use a more French-like accent, as noted by Ziamari (2008), this would contribute to the perception that they

use more French lexical items despite their similar numeric frequency to male use. The question of pronunciation is beyond the scope of the present inquiry, but should be investigated in the future through an instrumental analysis of speech production, using the current corpus or another, as well as a perception task in which listeners describe the amount of CS they note in a given person's speech.

6.3 Effect of proficiency on CS use

Proficiency in each language is often found to be a significant factor in the use of CS, as reported in chapter 3 above. As native speakers of Moroccan Arabic, residing in Morocco, all participants are highly proficient in their dialect, but their proficiency in French varies. All participants began learning French by age 8 at the latest, with the average age 7.1 years old at the time of the first formal exposure to French learning. However, some participants attended private schools that placed a greater emphasis on French fluency, and thereby received a greater amount of formal instruction in French. Exposure to the language outside of the classroom also may differ greatly between individuals based on parental language preferences, family television viewing habits, and neighborhood. For these reasons, participants' proficiency in French was gauged via two separate measures: a short written grammar test and a task-based self-report. Both of these measures are described in more detail in Chapter 5 above. Two participants (P10301, P10302) did not complete the written grammar test and are not included in the analysis for this reason. The French proficiency scores of the other participants are given in Table 6.18 below.

Table 6.18 Proficiency scores of participants included in extra-linguistic analysis

Dyad	Speaker	French grammar score	French self-rated proficiency
1	P10231	12	3.9
1	P10232	8	3.6
2	P10261	14	4.7
2	P10262	18	3.3
3	P10281	13	3.3
3	P10282	19	4.3
4	P10301	--	3.5
4	P10302	--	3.2
5	P11021	10	4.3
5	P11022	10	2.2
6	P11051	17	4.4
6	P11052	18	4.3
7	P11131	11	3.1
7	P11132	7	2.5
8	P11201	12	3.0
8	P11202	10	2.9
9	P11213	15	4.6
9	P11214	13	4.7
10	P11221	14	4.4
10	P11222	22	4.6
11	P11223	21	4.8
11	P11224	17	4.6
12	P11225	16	4.2
12	P11226	8	3.5
13	P11271	16	3.0
13	P11272	12	2.6
14	P11291	10	3.8
14	P11292	15	2.0
15	P11293	15	4.4
15	P11294	9	3.1
16	P12031	10	2.1
16	P12032	7	3.5

6.3.1 FRENCH PROFICIENCY and rate of French

The possibility that FRENCH PROFICIENCY may affect the *rate* of French is indicated by Sayahi's (2011a) results. He found that males with a high school education used less French than males with a university education. Using level of education as a proxy for French proficiency due to the fact that university studies are carried out almost entirely in French in Tunisia, he proposes that this difference may be due to the difference in proficiency levels between speakers. This question has not been explored in other studies on Arabic-French CS. Bentahila and Davies (1995) assume that their older generation is more proficient in French, but there is no evidence that this group uses more French than the younger. The authors report a similar number of switch types used by the two groups, but these must not be confused with rates of French use due to the vast difference in the recording lengths for each generation. The data from the older generation is drawn from approximately six hours of conversation and contains 757 switches, while that of the younger generation is drawn from approximately 2.5 hours of conversation and contains 788 switches. The younger generation clearly has a higher rate of switches per hour, but full clauses, the constituent type preferred by the older generation, are inherently longer than the noun phrases used most often by the younger generation. Without a count of the words or constituents used in Arabic, it is impossible to reliably compare the rate of French used by each group. With the results from these two studies in mind, it is cautiously predicted that speakers with a higher proficiency in French will use more French lexical items in their Arabic.

6.3.1.1 FRENCH PROFICIENCY and rate of French in Instant Messaging

PROFICIENCY as a possible explanatory factor was first examined in reference to the written chat production. A cursory visual inspection of the data as depicted in Figure 6.1 reveals that no relationship exists between proficiency and French rate of use in IM. I

analyzed these results using a linear mixed effects analysis of the relationship between rate of French in writing and FRENCH PROFICIENCY in R (R Core Team 2012) using lme4 (Bates, Maechler & Bolker 2012). The fixed effect in the model was FRENCH PROFICIENCY, with speaker pair as a random effect. Rate of French in writing was the dependent variable. Visual inspection of the data did not indicate any violations of normality. A likelihood ratio test of the model including FRENCH PROFICIENCY and the model without this factor was used to obtain a P-value. This result confirms that FRENCH PROFICIENCY does not impact the use of French in the written data (p= 0.6988).

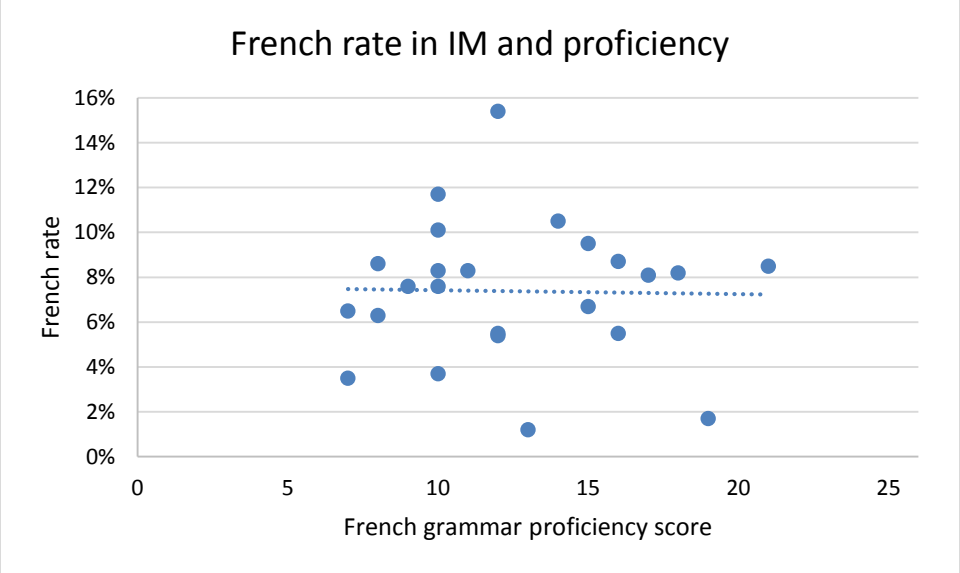

Figure 6.1 Rate of French in written task and French proficiency

Self-ratings of French proficiency were also obtained from participants as part of the background survey administered after the completion of all tasks. As described in the methodology section, participants rated their ability to perform a range of production and comprehension tasks based on real-world situations. The self-ratings are seen with the French rate in Figure 6.2 below.

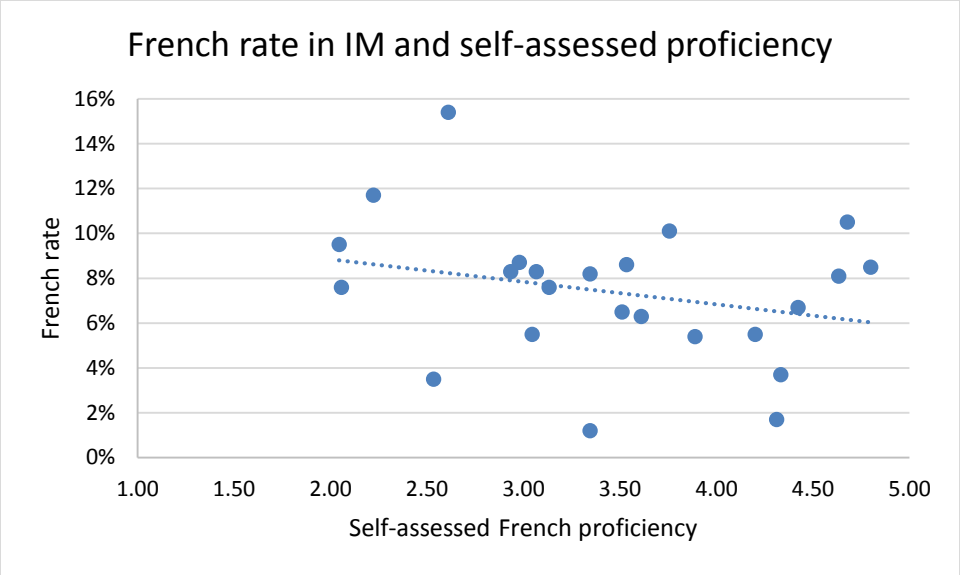

Figure 6.2 Rate of French in written task and self-assessed proficiency

I analyzed these results using a linear mixed effects analysis of the relationship between rate of French in writing and self-assessed French Proficiency in R (R Core Team 2012) using lme4 (Bates, Maechler & Bolker 2012). The fixed effect in the model was self-assessed FRENCH PROFICIENCY, with speaker pair as a random effect. Rate of French in speaking was the dependent variable. Visual inspection of the data did not indicate any violations of normality after the data was transformed. A likelihood ratio test of the model including self-assessed FRENCH PROFICIENCY and the model without this factor was used to obtain a P-value. This result indicates that the effect of self-assessed French Proficiency is not significant in the written data (p= 0.1423).

6.3.1.2 FRENCH PROFICIENCY and rate of French in spoken conversation

The spoken data was similarly examined for a potential effect of proficiency on the rate of French use. The rate of French in speaking is plotted with the grammatical proficiency scores in Figure 6.3 below. The French rate shown in this figure is of the

transformed data and for this reason does not correspond to the raw percent of French used. However, this figure can be interpreted in a similar manner as those above; speakers that have a transformed French rate closer to 0, and thus the top of the figure, have the highest rates of French use.

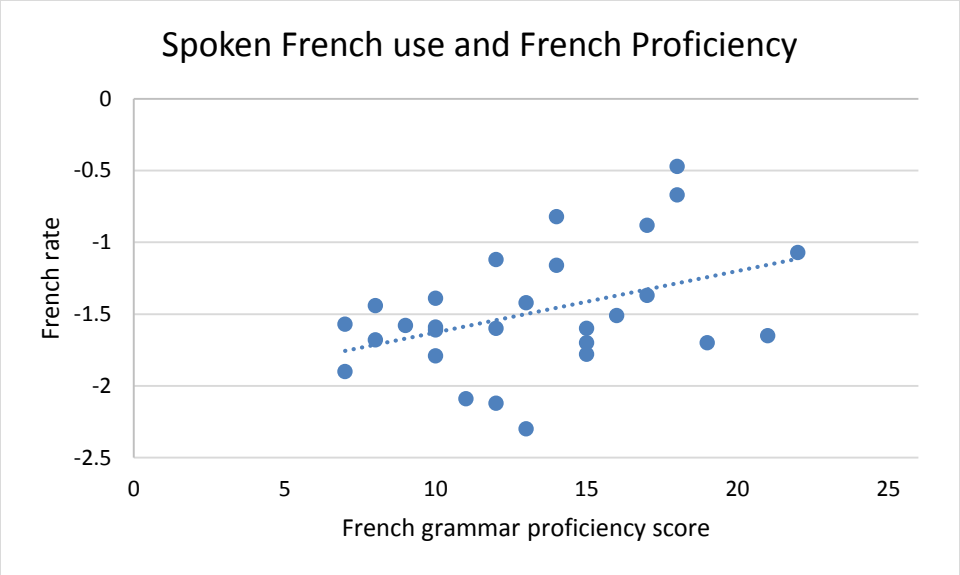

Figure 6.3 Rate of French in spoken task and French Proficiency

I analyzed these results using a linear mixed effects analysis of the relationship between rate of French in speaking and French Proficiency in R (R Core Team 2012) using lme4 (Bates, Maechler & Bolker 2012). The fixed effect in the model was French Proficiency, with speaker pair as a random effect. Rate of French in speaking was the dependent variable. Visual inspection of the data did not indicate any violations of normality after the data was transformed. A likelihood ratio test of the model including French Proficiency and the model without this factor was used to obtain a P-value. This result reveals that French Proficiency does not impact the use of French in the spoken data (p= 0.7551) despite the appearance of a trend in Figure 6.3.

The potential effect of self-assessed French proficiency was also analyzed within the spoken data. A chart showing the self-assessed score of all participants is available in Figure 6.4 below with the rate of French lexical items used in their speech.

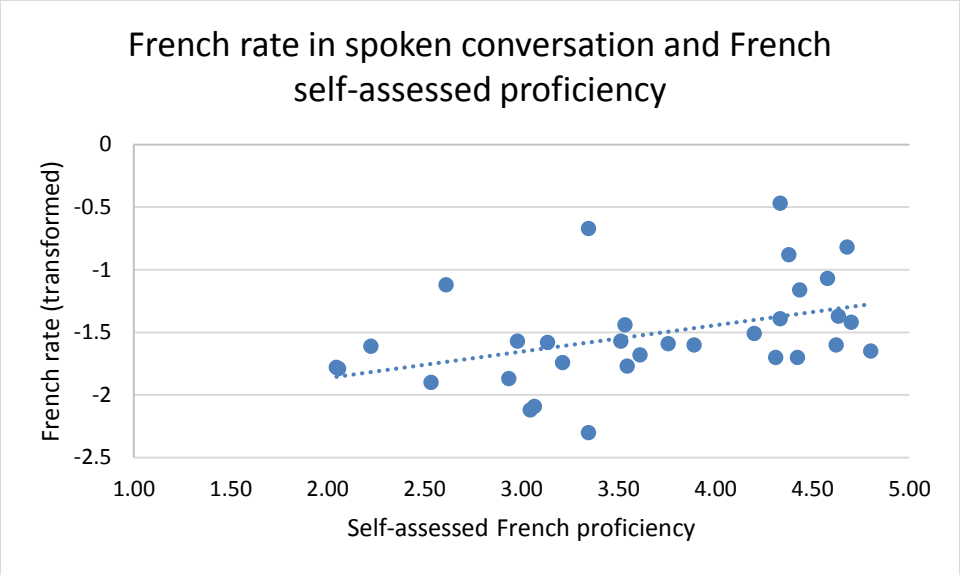

Figure 6.4 Rate of French in spoken task and self-assessed French proficiency

These results were also analyzed using a linear mixed effects analysis of the relationship between rate of French in speaking and self-assessed FRENCH PROFICIENCY in R (R Core Team 2012) using lme4 (Bates, Maechler & Bolker 2012). The fixed effect in the model was self-assessed French Proficiency, with speaker pair as a random effect. Rate of French in speaking was the dependent variable. Visual inspection of the data did not indicate any violations of normality after the data was transformed. A likelihood ratio test of the model including FRENCH PROFICIENCY and the model without this factor was used to obtain a P-value. This result reveals that self-assessed French Proficiency trends toward significance, but does not reach it, in the spoken data (p= 0.07688).

6.3.1.3 Discussion

While proficiency has often been found to be related to the use of CS, this result is not supported in the current results. Sayahi's (2011a) data suggests an effect for proficiency based on level of education, but as he did not assess proficiency level or perform a statistical analysis, it is impossible to verify if the trend he observed in his data is significant. None of the participants performed at ceiling on the paper-based grammar test, indicating that this test was appropriately challenging for all participants. One possibility is that certain French lexical items not previously attested as established borrowings in MCA are so widely used that even speakers with limited French know and employ these phrases.

The self-ratings indicate a trend toward a relationship between this rating and use of French in speaking, but not in written production. Such ratings provide a global measure as they are not strict measures of linguistic proficiency but instead may implicitly include attitude toward the language or other sub-conscious aspects of language use that are difficult to identify. The results here indicate that self-ratings should be considered in future studies in order to determine whether the observed trends are found in other data sets.

6.3.2 FRENCH PROFICIENCY and constituents switched

With the wide range of proficiency in French, it is hard to gauge how this factor might affect the constituents produced in French. However, Poplack (1980) implicates proficiency in how individuals use CS, particularly in the distinction between inter- and intra-sentential forms. PROFICIENCY is considered here, but as Sayahi (2011a) did not find this to be a decisive factor for his participants, it is not expected to be here, either. To explore the effect of PROFICIENCY on the type of constituents that are switched by these speakers, the participants were divided into four groups that correspond to four quartiles. This means that the division between the med-low and med-high groups is the median

proficiency score, with participants divided into roughly equal groups, based again on proficiency score, above and below the median. This was completed separately for the each communication modality as the corpora overlap for only twelve participant pairs, with four pairs added to the spoken analysis.

6.3.2.1 FRENCH PROFICIENCY and French constituents in Instant Messaging

For the IM data, the median French proficiency score is 12, and the proficiency groups are seen in Table 6.19.

Table 6.19 Proficiency ranges for groupings in written data

	Proficiency range	# of participants
Low	7-9	5
Med-low	10-12	9
Med-high	13-15	6
High	16+	4

When the written data is considered by proficiency quartile, there are very few patterns that stand out, as seen in Table 6.20. Differences exist between proficiency groups in use

Table 6.20 French constituents in written corpus by proficiency group

Constituent	**Low**	**Med-Low**	**Med-High**	**High**	**Full Group**
Noun/Noun Phrase	23.1%	39.1%	28.0%	36.7%	33.6%
Oui/Non	20.5%	23.6%	22.7%	19.3%	21.7%
Conjunction	12.8%	6.8%	5.3%	11.9%	9.0%
Complementizer Phrase	12.8%	8.7%	8.0%	3.7%	8.0%
Determiner Phase	6.4%	6.8%	9.3%	10.1%	8.0%
Verb/Verb Phrase	9.0%	3.1%	10.7%	6.4%	6.4%
Adjective/Adj Phrase	3.8%	4.3%	5.3%	5.5%	4.7%
Interjection	10.3%	3.1%	4.0%	3.7%	4.7%
Adverb/Adv Phrase	1.3%	3.7%	2.7%	1.8%	2.6%
Preposition/Prep Phrase	0.0%	0.6%	4.0%	0.0%	0.9%
Number Phrase	0.0%	0.0%	0.0%	0.9%	0.2%

of Nouns/Noun Phrases, but there is neither a clear increase or decrease with proficiency as the Medium-low proficiency group displays the highest use of Nouns. Many constituent types are used at similar rates across the board, such as Oui/Non. However, a few notable trends emerge: first, full Determiner Phrases increase with higher proficiency levels. At the same time, complementizer phrases gradually decrease with higher proficiency. Interjections also sharply drop off from the lowest proficiency speakers in the first quartile using the most, while similar use is found among all higher proficiency levels. The differences observed in these constituent types across proficiency levels may indicate that proficiency in French may impact the type of French constituents used by speakers, even though no difference is found for rate.

6.3.2.2 FRENCH PROFICIENCY and constituents switched in spoken conversation

The participants were again split into four proficiency groups based on quartiles, as described in §6.3.2.1 above, for analysis for the French constituents used in the spoken data. However, as the spoken corpus contains a larger number of participants than the written, the proficiency groups vary slightly, as shown in Table 6.21.

Table 6.21 Proficiency ranges for groupings in spoken data

	Proficiency range	# of Participants
Low	7-9	5
Med-low	10-13	11
Med-high	14-16	7
High	17+	7

As with the written data, trends emerge for constituent types across proficiency groups in the spoken data, seen in Table 6.22. In the spoken data, the rate of Noun/Noun Phrases decreases with an increase of proficiency, but remains the largest proportion of French constituents across proficiency groups. As a contrast, use of Determiner Phrases increases with proficiency and is similarly always the second largest proportion of

constituents in the data. Two other constituent types show an increase in use with higher proficiency: Adjective/Adjective Phrases and Oui/Non.

Table 6.22 French constituents in spoken corpus by proficiency group

Constituent	Low	Med-Low	Med-High	High	Full Group
Noun/Noun Phrase	58.4%	51.7%	50.9%	35.9%	45.3%
Determiner Phase	28.6%	28.3%	29.5%	32.0%	30.3%
Verb/Verb Phrase	2.6%	7.8%	4.3%	3.4%	4.5%
Adjective/Adj Phrase	1.3%	3.9%	4.1%	5.5%	4.4%
Complementizer Phrase	1.3%	0.0%	1.9%	7.5%	3.8%
Number/Num Phrase	2.6%	1.3%	1.9%	4.9%	3.1%
Oui/Non	1.3%	1.7%	2.2%	3.2%	2.5%
Adverb/Adv Phrase	0.0%	2.6%	1.6%	2.8%	2.2%
Conjunction	2.6%	1.7%	1.1%	1.2%	1.4%
Preposition/Prep Phrase	0.0%	0.4%	0.8%	2.2%	1.3%
Interjection	1.3%	0.4%	0.5%	0.4%	0.5%
Definite Article	0.0%	0.0%	1.1%	0.2%	0.4%
Quantifier Phrase	0.0%	0.0%	0.0%	0.4%	0.2%
Complementizer	0.0%	0.0%	0.0%	0.2%	0.1%

6.3.2.3 Discussion

FRENCH PROFICIENCY relates to different trends in each communication modality with little overlap. In both modalities, the use of Determiner Phrases increases with higher French proficiency. However, in IM chat the use of Complementizer Phrases and Interjections decreases with proficiency, as would be expected from Poplack's (1980) suggestion that such extra-sentential elements are easier to incorporate into CS. These extra-sentential constituent types do not show any trends for FRENCH PROFICIENCY in the spoken data, indicating that modality plays a role in their use among these speakers and not proficiency alone. Within the spoken data, different trends emerge as the use of

Noun/Noun Phrases decreases with higher proficiency as well as the three other constituent types noted above.

6.4 Effect of Language Attitude on CS use

MacSwan (1997) notes that he used code-switching data only from informants who approve of the practice as attitude can notably impact a speaker's use of each language. The differences in grammaticality judgments based on attitude found by Anderson (2006) support this possibility. For this reason, a basic measure of attitude toward French was included in the current analysis based on respondents' multiple-choice answers to two prompts included at the end of the background questionnaire: "What is/are the most practical language(s) for everyday life" and "I prefer…" For each question, respondents were given five options: Moroccan Arabic, French, Mix of Moroccan Arabic and French, Standard Arabic, or Other (with a blank to specify the language.) If *French* or *Mix of Moroccan Arabic and French* was included in these questions, a participant was given one point, for a possible maximum of two points. While LANGUAGE ATTITUDE is a complex and multi-faceted phenomenon, this basic measure was employed to conduct an initial inquiry related to this factor.

6.4.1 Language Attitude and rate of French

Previous studies on Arabic-French CS have not examined the potential impacts of LANGUAGE ATTITUDE directly on the use of each language. However, Chakrani (2010) provides reported use data on both attitude and use. He finds that Moroccan young adult speakers in the Marrakesh area report using French in a broad range of settings, including within the home, especially among individuals of the middle and upper classes (145). At

the same time, he discovers a positive attitude toward French in the same groups. For this reason, it was predicted that a positive attitude will lead to greater use of French and CS.

6.4.1.1 LANGUAGE ATTITUDE and rate of French in Instant Messaging

In the written data, all participants use a similar amount of French, regardless of language attitude, with those who fall into the middle range of attitude using the fewest French lexical items. Table 6.23 gives the average rate of French lexical items by participants, divided by attitude rating.

Table 6.23 Rate of French and language attitude in written production

Attitude rating	Average French rate	# of Participants
0	7.3%	9
1	6.4%	11
2	9.7%	4
Overall average rate	7.4%	24

Unsurprisingly, a statistical analysis shows no relationship between LANGUAGE ATTITUDE and rate of French in IM. A linear mixed effects analysis was carried out to test this relationship in R (R Core Team 2012) using lme4 (Bates, Maechler & Bolker 2012). The fixed effect in the model was attitude rating, with speaker pair as a random effect. Rate of French in writing was the dependent variable. Visual inspection of the data did not indicate any violations of normality. A likelihood ratio test of the model including LANGUAGE ATTITUDE and the model without this factor was used to obtain a P-value. This result reveals that positive attitude toward French does not impact the use of French in the spoken data (p= 0.329).

6.4.1.2 LANGUAGE ATTITUDE and rate of French in spoken conversation

The spoken data shows a different picture from the written data, as seen by the differential rates of French use in Table 6.24. Here, it is clear that rate of French use

increases greatly with positive attitude toward French and the highest approval rate coincides with a French rate more than double the average.

Table 6.24 Rate of French and language attitude in spoken production

Attitude rating	Average French rate	# of Participants
0	1.8%	11
1	4.1%	14
2	12.9%	7
Overall average rate	**5.1%**	**32**

Once again, a linear mixed effects analysis was carried out to test the relationship between LANGUAGE ATTITUDE and Rate of French in spoken language using R (R Core Team 2012) and the lme4 package (Bates, Maechler & Bolker 2012). The fixed effect in the model was again attitude rating, with speaker pair as a random effect. Rate of French in speaking was the dependent variable. Visual inspection of the data did not indicate any violations of normality after data was transformed. A likelihood ratio test of the model including LANGUAGE ATTITUDE and the model without this factor was used to obtain a P-value. This result confirms that the vast difference observed in Table 6.24 is highly significant (p = .00195).

6.4.1.3 Discussion

The lack of effect for attitude in writing may indicate that the higher use of French is a sign of the emerging norm in this modality for all speakers, a possibility that is supported by the higher average use of French in written communication. On the contrary, the use of French in spoken conversation is highly dependent on an individual's attitude toward the French language and the practice of CS. This result provides speaker-level support to Chakrani's (2010) group-level finding that the classes of speakers with positive attitudes toward French and CS are more likely to use French in their conversations.

6.4.2 LANGUAGE ATTITUDE and constituents switched

As the type of constituent switched showed differences for SEX and FRENCH PROFICIENCY, it was also modeled as a function of LANGUAGE ATTITUDE. As this type of analysis has not been completed in prior research, predictions regarding the types of constituents switched by LANGUAGE ATTITUDE are difficult to make. However, general trends in the types of constituents used most broadly in CS by this group would suggest that Noun/Noun Phrases will be used most often by speakers of all groups, followed by Complementizer Phrases, Oui/Non and Interjections.

6.4.2.1 LANGUAGE ATTITUDE and constituents switched in Instant Messaging

Table 6.25 displays the proportion of French constituents of a given type out of all the French constituents used in the IM data by the target subgroup. The subgroups in this table represent participants with the same attitude rating, with 2 being the most positive attitude toward French and CS. The majority of constituent types are quite similar across attitude groups, or do not show a trend that is directly related to attitude. One distinction is visible in the use of Nouns and Noun phrases; positive attitude toward French may be reflected in a higher proportion of other constituent types as the use of Noun/Noun Phrases is highest for the group with the least positive attitudes toward French. The use of Complementizer Phrases is also very different between groups, with speakers who approve of the use of French and CS using the highest proportion of full clauses in their writing. The final distinction of note is the use of adverbs that increases with positive attitude toward French and CS; this indicates a positive relationship, but at the same time the low level of Adverbs in the written corpus makes it necessary to investigate whether the same relationship might be found in a larger corpus.

Table 6.25 Language attitude ratings and French constituents used in writing

French Constituent	Attitude rating			Full Group
	0	1	2	
Noun/Noun Phrase	43.9%	27.2%	26.1%	32.5%
Oui/Non	17.6%	23.8%	19.1%	20.1%
Conjunction	4.9%	9.4%	7.5%	7.3%
Complementizer Phrase	4.9%	7.4%	20.6%	10.9%
Determiner Phrase	9.8%	11.4%	6.5%	9.2%
Verb	7.3%	7.4%	7.5%	7.4%
Adjective	5.4%	4.5%	4.0%	4.6%
Interjection	3.9%	4.5%	2.5%	3.6%
Adverb	1.0%	3.5%	4.0%	2.8%
Preposition/Prep Phrase	1.5%	1.0%	1.5%	1.3%
Number Phrase	0.0%	0.0%	0.5%	0.2%

6.4.2.2 LANGUAGE ATTITUDE and constituents switched in spoken Conversation

Table 6.26 gives the proportion of French constituents of a given type out of all the French constituents in the data used by the target subgroup. The subgroups in this table represent participants with the same attitude rating, with 2 being the most positive attitude toward French and CS. Many of the trends noted in the written data by language attitude ratings are seen again in the spoken data: the proportion of Nouns/Noun Phrases used declines with higher approval of French and CS. At the same time, several constituents display an increase in proportion of French used with positive attitude toward French and CS: Number phrases increase substantially, as does use of full clauses. While minor constituent types throughout the data, use of prepositions and preposition phrases and adverbs/adverb phrases is higher among the participants with the most positive attitude toward CS, indicating that this group uses a wider range of constituents than the other speakers.

Table 6.26 Language Attitude and French constituents used in speech

French Constituent	Attitude rating			Full group
	0	1	2	
Noun/Noun Phrase	56.2%	48.9%	39.6%	46.0%
Determiner Phrase	28.3%	30.6%	29.6%	29.7%
Adjective/Adj Phrase	2.3%	5.6%	4.1%	4.2%
Complementizer Phrase	1.1%	0.6%	6.9%	3.6%
Number/Num Phrase	3.4%	7.3%	3.0%	4.5%
Verb/Verb Phrase	3.4%	4.2%	2.3%	3.2%
Oui/Non	0.8%	1.9%	4.8%	3.1%
Adverb/Adv Phrase	1.5%	1.5%	4.4%	2.8%
Preposition/Prep Phrase	3.0%	0.9%	3.0%	2.3%
Conjunction	1.9%	1.1%	1.1%	1.2%
Interjection	0.4%	0.4%	0.6%	0.5%
Definite Article	0.8%	0.4%	0.3%	0.4%
Quantifier Phrase	0.0%	0.0%	0.3%	0.1%
Complementizer	0.0%	0.0%	0.2%	0.1%

6.4.2.3 Discussion

These findings regarding the role of LANGUAGE ATTITUDE in rate and constituents employed in code-switching show a robust effect for LANGUAGE ATTITUDE, a factor that has been implicated in grammaticality judgments but not investigated in usage. Attitude seems to play an important role in the use of CS in spoken language, where it is significantly related to the rate of French employed. In fact, the effect of this factor is the strongest of any of the factors analyzed in this data set and for this reason should be investigated in other contact settings to identify whether it may impact CS across language pairings. This effect is not seen for written language, which is again seen to have more consistent French use between speaker groups. While many constituent types display differences by SEX and PROFICIENCY, very few constituents suggest a trend for LANGUAGE ATTITUDE. From these observations, it seems that factors such as SEX and PROFICIENCY

impact the *types* of constituents that a speaker uses, while LANGUAGE ATTITUDE plays a role in *how often* the speaker uses those constituents.

6.5 CONCLUSION: RATE OF FRENCH AND CONSTITUENTS USED

This analysis reveals important discoveries within the traditional modality of communication in CS studies, spoken conversation, as well as in written IM. The research design permits an unparalleled opportunity to directly compare the spoken and written communication of the same individuals in order to pinpoint the differences that may arise as a result of communication modality. These differences exist in the *rate* of French insertions used, a factor often commented on but rarely quantified, as well as differences in the *types* of syntactic constituents employed. This approach reveals distinct uses of French lexical items that are characteristic of each modality in addition to divergent effects of the extra-linguistic factors by modality.

Comparison of the current data with past results is tentative due to the reporting methods of each study, but limited comparisons can be made. While it is impossible to compare rate of French lexical items across studies, some studies do report the number of instances of code-switching in their data as well as the hours recorded, giving a rough estimate of amount of CS. Sayahi's (2011a) data contains 1,721 instances of CS in a three-hour spoken sample, while the current data contains 1,370 uses of French in 4.25 hours. While the number of switches per hour is less than those found in Sayahi's Tunisian data, it remains vastly higher than the 1,835 instances of CS culled from the 66 hours of sociolinguistic interviews among Spanish–English bilinguals, analyzed by Poplack (1980). Consideration of the proportion of switch types again highlights differences between the current corpus and past data as the spoken data analyzed here contains a very high

proportion of nominal structures compared to previous studies, as noted in the comparison between past studies and the current results given in Table 6.11 above.

Analysis of the spoken data in comparison with past studies alone points to important distinctions between the use of CS by young Moroccans today and other populations. The current data contains a lower rate of CS than the Tunisian data collected by Sayahi. The participants also use a much higher rate of nominal structures than indicated in previous research. While the current data does not suggest a correlation between use of nominal structures and proficiency, or allow a statistical analysis of this possibility, it may be that this factor is important across groups.

In considering communication modality on its own, two primary distinctions stand out. First, the participants as a group use significantly more French in writing than they do in speech. This is hypothesized to relate to the CMC environment and, specifically, to the setting of Facebook for the chat as this particular social networking service is commonly found to be associated with French in Morocco. Second, the French constituents in writing differ substantially from those in speech and indicate that it is the greater use of certain constituents, and not simply an overall increase in use of French in this modality that creates the observed distinction. This finding suggests that certain linguistic practices have become associated with the modality of communication by the current participants and highlights the need to use a model of CS that takes modality into account. The lower rates of Oui/Non production in the spoken conversations, despite their common use by the majority of speakers in writing, is the strongest evidence of these differing practices. The results also indicate differences in extra-linguistic factors based in part on modality of communication, a result that could not have been predicted due to the paucity of existing data on this matter.

SEX has been often implicated as relevant in the use of French for speakers of Arabic and French, but the data reveals that participants of both sexes use a similar amount of French in spoken conversation, against typical sex-based stereotypes. However, the written data reveals that male participants use more French in this modality, a contradiction of the common stereotype. The examination of French constituents used in speech may hint at why females may be perceived to use more French. The higher use of extrasentential elements by female speakers, such as interjections and complementizer phrases, may be more salient to listeners and increase the popular association of females with French. Males use a much higher proportion of nouns and noun phrases in writing than females, which may indicate that the subjects they discuss are technology, education, or others that are typically associated with the use of French. Combined with Ziamari's note that female speakers use a pronunciation of French words that is more faithful to the source language, we recommend investigation of the salience of constituents present in CS and the pronunciation of these constituents in regard to perceived level of CS participation.

Proficiency in French is shown not to contribute significantly to the variance in the rate and type of CS manifested among Moroccan young adults despite the fact that this factor has been considered an important factor in CS (Poplack 1980 among others). A statistical analysis shows that FRENCH PROFICIENCY, whether measured through a basic grammar test or self-assessed, does not impact the rate of French employed by the current participants. The lack of effect found for this variable, a contradiction of many past studies, may be due to the current methodology or to the target group itself. The methodology of this study used two measures of proficiency. The first measure, a traditional pen-and-paper based test of grammatical accuracy, may have underestimated the proficiency of speakers who are accustomed to using French in settings in which grammar is less important than interpersonal communication: on internet forums, through watching TV, in interactions

with foreigners, or in educational settings where code-switching is common. A verbal proficiency measure may provide more informative results for these speakers.

The scores on the self-assessed proficiency also have no relationship to the rate of French used among participants. However, the latter results did trend toward significance in the written modality; this result calls for future work to measure proficiency in a manner that takes into account how the speakers might commonly use the target language. On the other hand, it may be that CS practices within the target community are widely shared, as all young adults have extended exposure to the language through education and media, thus rendering the factor of proficiency less relevant than it may be in other contexts in which exposure to a contact language is less common. Yet another possibility is that the current participants represent a limited sample of the full population and that an effect for proficiency may be visible in a sample that incorporates a broader subsection of the population.

Finally, LANGUAGE ATTITUDE was considered as an exploratory factor and significant results are revealed in a surprising way; this factor has no effect in informal written CMC, but is highly significant in spoken conversation. This result may indicate that IM language practices may be similar among speakers of different backgrounds as they conform to emerging norms within the community. A high use of French, including Oui/Non in written MCA seems to be the unmarked practice in IM, while its use in spoken conversation is rare and related to LANGUAGE ATTITUDE. The measure of LANGUAGE ATTITUDE shows the strongest relationship with rate of French of all the extra-linguistic factors examined and should be examined in future research. The strength of the results for this measure indicates that it is not a factor that can be ignored in the search for how and why speakers employ two languages within the same conversational turn.

One observation remains constant throughout these results: all participants use a high proportion of French nominal structures when compared to all French constituents used. As nouns are often found to be the most commonly borrowed part of speech, or the easiest to borrow, a logical question is whether this apparent use of French nouns is in fact CS or borrowing. It is for this reason that Chapter 7 explores the use of French nouns and attempts to distinguish between lexical items of French origin that might now be best considered fully part of MCA and those that are safely considered to be instances of CS.

Chapter 7: Common French-origin nouns

7.1 Introduction

The analysis of the CS data presented in the previous chapter indicates that the majority of French constituents used in the current corpus are nominal structures, defined as nouns, noun phrases and determiner phrases. Many studies note that nouns are the most common part of speech to be borrowed, leading us to question whether the high rate of nouns observed in the present data are in fact instances of code-switching (see Muysken 2000 for a summary of proposals on the distinction between CS and borrowing). There is reason to suspect that nouns that appear in Determiner Phrases could also be borrowings; Heath (1989) notes that certain French lexical items, such as *la gare*, 'the train station,' have been borrowed into MCA with their definite article. For this reason, I completed a further analysis of the most common French-origin nouns present in the corpus.

While the primary analysis assumes that all French-origin nouns absent from bilingual dictionaries and Heath's study of borrowing in MCA are code-switches, it is not clear that these sources document the lexicon of MCA well enough to be a reliable basis for this distinction. Two of the three MCA bilingual dictionaries pre-date many of the technologies that are used regularly in Morocco today, as they are without substantial updates since their original publication in 1959 and 1966; Heath's (1989) study was similarly completed before widespread use of cell phones, computers, and the internet, among other things. The one recent dictionary (Aguadé & Benyahia 2013) contains few lexical items related to technology or education, listing *ḅoṛṭabl*, 'cell phone,' but no entry for 'internet.' A reexamination of borrowings into MCA is clearly warranted. While this issue cannot be fully addressed in the current corpus, the high rate of nominal structures allows us to examine the use of nouns in the data for evidence of borrowing.

The distinction between code-switching and borrowing, discussed in §3.1, is again relevant here. Even those who consider that delimiting the two is important to language theory admit that doing so is not simple. Poplack et al (1988) define borrowing as a donor language lexical item that has been morphologically, syntactically and phonologically integrated into the recipient language. CS for these authors is then the use of one or more lexical items that is morphologically and syntactically grammatical in one language. Muysken (2000;73) interprets this distinction to mean that CS is at or above the level of the word, while borrowing is at or below the level of the word. This leaves the use of single lexical items difficult to classify. Muysken further notes that the notion of *listedness* is relevant in discussions of CS and borrowing, as multi-word phrases of donor language may become borrowed and accepted in a given community as the most common way to express a certain notion, while single words may be part of a code-switch. Muysken states that it is the listedness of a lexical item, or the notion within a speech community that the lexical item is part of their language, that helps to identify whether the use of donor language material is CS or borrowing, regardless of the length.

In order to examine the distinction between CS and borrowing experimentally, Poplack and Sankoff (1984) elicited nouns from Puerto Rican Spanish-English bilinguals in New York. They analyzed the nouns gathered in reference to six factors in an effort to determine which could be confidently identified as borrowings. The factors they selected to identify borrowings are:

1. Displacement of the recipient language word by that of the donor language
2. Displacement of donor language synonyms by a single donor language lexical item
3. Degree of phonological and morphological integration into the recipient language

4. Consistency of gender assignment
5. Number of speakers using a lexical item ('diffusion' in my terms)
6. Accepted by speakers (1984:103-104)

Phonology is excluded from the current analysis due to the current focus on structure and the tendency for French phonemes to be borrowed into MCA. Gender assignment is similarly left aside as gender is not expressed on Arabic articles, leaving few of the attested nouns marked for gender. This study accesses acceptability by assuming that the elicited forms for a referent are acceptable to a speaker, and thus acceptable within a given community when widely used by speakers.

The size of the corpus created for this study provides an opportunity to analyze these nouns in greater detail in an effort to determine whether any nouns can confidently be labeled as code-switching, or if they have instead become a part of the Moroccan Arabic lexicon despite their absence from Moroccan Arabic dictionaries. This question is particularly relevant as many of the French-origin nouns in the corpus are related to the domains of education and technology, both traditionally associated with French in Morocco. In addition, in listening to the files and reviewing the conversation transcripts, we get the impression that the logical Arabic translation equivalents are not used for many of these nouns. For these reasons, I explore the most widely used French-origin nouns in the corpus in order to understand if they are rightfully categorized as CS, or may in fact be better described as borrowings.

In undertaking this analysis, we enter into the controversial topic of the distinction between borrowing and code-switching discussed in Chapter 3. I take a broad approach to this problem by examining each lemma, which Crystal (1997) defines as "an abstract representation, subsuming all the formal lexical variations which may apply: the verb walk, for example, subsumes walking, walks and walked or citation form of the word." For

nouns, this means that any plural or gender information is removed, and abbreviations can be treated with full forms. In this way French lemma 'professeur' may be realized as 'le/la prof,' 'le professeur,' 'la professeure' as well as the plural form of each.[23] In this way I do not attempt to label individual tokens as belonging to either of these contact phenomena, but focus on the several aspects of the use of each lemma throughout the corpus.

After all tokens of a lemma are considered, I indicate whether there is evidence that the lemma has become a part of MCA lexicon, and thus could have the status of a borrowing, or remains an instance of code-switching from French. All French-origin lemmas are assumed to belong only to French, and therefore represent instances of code-switching, unless the data reveals that they may now be considered borrowings. By identifying the lemma as a borrowing or code-switch, we create a more complete picture of modern MCA and the borrowing routines in this contact setting. While the current corpus is too small to be fully representative of MCA, certain topics recur in the conversations of many participant pairs and the lexical items related to these conversational topics may thus shed light on the current lexicon used by young adults in Morocco. The findings of this analysis can then be compared to a broader speaker group to confirm or correct these initial conclusions.

7.2 METHODOLOGY

The goal of this chapter is to examine the French nouns occurring in the corpus to better understand how a noun lemma is incorporated into MCA. This analysis was applied to any and all of the nouns that are used by five or more speakers in the corpus. This level of diffusion is arbitrary, but the cut-off at a minimum of 5 speakers was chosen as that

[23] Knowles and Zuraidah (2004) point out the inadequacy of this definition for Arabic lemmas, but as only French nouns were lemmatized, this issue is not further addressed here.

number represents more than ten percent of the speakers in the corpus. Four speakers would also meet this standard, but in this case it would be possible for all four speakers to be found in two conversation pairs; five was preferred in order to eliminate this possibility and thereby assure use in at least 3 of the 18 participant dyads. This is a relatively high threshold as a measure of diffusion when compared to similar analyses, such as that used by Poplack, Sankoff and Miller (1988), which set the highest level of diffusion at ten of 120 speakers, or 8.3% of the speakers in the Ottawa-Hull corpus. Using our threshold that a token must be present in the output of 5 the 38 speakers, the corpus yields a total of 30 French-origin nouns for the purposes of this analysis.

The French-origin nouns with the highest diffusion rate, were each analyzed using four metrics:

1) Frequency of French lemma, measured as the raw token count in the data

2) Diffusion of the lemma, measured as the number of speakers who use it

3) Observed syntax and morphology of the token

4) Frequency of possible translation equivalents in Arabic (MCA and Standard)

These metrics are similar to the factors analyzed by Poplack and Sankoff (1984), but exclude the factors of phonology and gender assignment identified by these researchers for the reasons noted above. Acceptability is assumed here to be reflected in a high level of use in the corpus. Lexical items that are used by few speakers in the corpus may also be acceptable and simply not attested in the current corpus; a larger corpus would help to clarify this possibility. Once the frequency of the Arabic translation equivalents was identified within the corpus, this information was combined with the frequency of French lemmas to give a French 'success rate,' (Zenner et al. 2012) calculated as the percent of times the referent is expressed in French out of all tokens of the lemma (λ) used to express the referent in the corpus, in other words:

$$\text{Success rate} = \frac{\textit{Token count of } \lambda \textit{ in French}}{\textit{Token count of } \lambda \textit{ in French} + \textit{Token count of } \lambda \textit{ in Arabic}} \times 100$$

French success rate is then similar to Poplack and Sankoff's measure of synonym displacement, where a higher success rate indicates a greater displacement of synonyms.

This multi-pronged approach incorporates several factors that may indicate the status of a lemma as either borrowed or as an instance of CS. No expectation was set for one of these factors to take priority over the others in determining which lexical items might now be part of Moroccan Arabic. Frequency and diffusion are numeric measures that are simple to observe. Observed morphosyntax is more complex as any article or adjective present with the target nouns must be considered; this process is further described in §7.4. Finally, the use of Arabic translation equivalents for the target lemmas drawn from both MSA and Moroccan Arabic are considered. It is hoped that the combination of these metrics will enable reliable identification of the target French-origin nouns as CS or borrowings.

7.3 NOUN MORPHOSYNTAX IN FRENCH AND MOROCCAN ARABIC

As the distinction between borrowing and code-switching is sometimes proposed on the basis of the morphosyntax of the target lexical items (Sankoff and Poplack 1981, Poplack, Sankoff and Miller 1988), it is critical to understand the structure of the determiner phrase (DP) in both Moroccan Arabic and French monolingual contexts. The structure of the DP between these languages is similar in the respect that both have a Determiner Phrase as the maximal structure that may reflect features of the noun. However, how features are expressed varies significantly between them.

7.3.1 Definiteness, gender and number

The traditional analysis of the morphosyntax of French and MCA states that each include a definite article, as seen in 7.1 and 7.2 respectively.[24]

7.1 a) La fille
The girl

b) Le garcon
The boy

c) Les filles
The girls

d) Les garçons
The boys

7.2 a) l-bənt
the=girl

b) d-dərri
the=boy

c) l-bnat
the=girls

d) d-drari
the=boys

Key differences are visible in these examples. First gender is expressed through the French definite article for singular animate and inanimate nouns, as seen in 7.1a and 7.1b. The definite article in MCA, like in other Arabic varieties, expresses only definiteness, seen in 7.2. While typically pronounced /l-/, the MCA definite article assimilates to initial coronal consonants, as in (7.2b) and (7.2d). Number is expressed through the definite article in French, visible in 7.1 c and 7.1d, but the gender feature is not overt in the plural. Although the French noun also carries a final <s> in writing, this plural marker is not pronounced,

[24] But see Turner (2014) for a convincing argument that the traditional assessment of definiteness in MCA should be revisited because the traditional definite article may be used in multiple indefinite contexts.

leaving only the article, phonetically [le] as opposed to singular [la] F.SG., or [lœ] M.SG., to convey the plural meaning in speech.[25] By contrast, number is marked by a change in vowel pattern in the Arabic root but not in the definite article. Many patterns of pluralization exist, but all patterns maintain the consonants of the root (i.e., the three to four letters that express the core meaning of a lexical item). The definite articles are typically distinct between the two languages. However, when a noun begins in French with a vowel, the final vowel of an article is elided thereby removing the information regarding gender and rendering the definite article identical to that of MCA (e.g., *l'école,* 'the school,' from feminine definite /la/).

The behavior of the French indefinite article is similar to that of the definite, but the indefinite feature is encoded in a substantially different way in MCA. The examples in 7.3 illustrate the French indefinite form of the DP.

7.3 a) Une fille
A girl

b) Un garcon
A boy

c) Des filles
Some girls

d) Des garçons
Some boys

Again, gender is encoded in singular indefinite articles, while there is a syncretic form of the plural indefinite article in French. This contrasts sharply with the MCA indefinite form; three indefinite structures are found in this variety. The first, a bare noun with no indefinite article, is shared with MSA and other dialects, as in 7.4.

[25] A small number of French nouns have a plural form that is phonetically distinct from the singular, such as 'cheval' and 'chevaux,' 'horse/s.' None of these nouns are present in our corpus.

7.4 a) bənt
A girl

b) dərri
A boy

c) bnat
Some girls

d) drari
Some boys

It is notable that while this structure is consistently included in descriptions of MCA (Marçais 1977, Caubet 1983, Youssi 1992, Harrell 1962, Maas 2011) authors rarely comment on its frequency. Those who do remark that it is rare and indicative of non-native speakers or of those who have extensively studied Standard Arabic (Harrell 1962, Chafik 1999).

Two other indefinite structures exist in MCA and are commonly used by speakers. The first common indefinite noun structure in 7.5 is unique to MCA

7.5 a) waħəd l-bənt
A girl

b) waħəd d-dərri
A boy

This structure uses *waħəd*, meaning 'one,' and has been described as a single indefinite article, /waħəd l-/ (Brustad 2000, Caubet 1983) or as a determiner that selects for a full DP (Harrell 1962, Youssi 1992). It is typically found with a singular noun and is often translated to mean 'a' in English, referring to a specific entity known to the speaker but not the addressee. The second indefinite structure utilizes the particle /ʃi/ as in 7.6 and is also common.

7.6) lqina **ši** bənt mʕ **ši** waħəd waqfin f **ši** qənt
found=1st.PL some girl with some person standing in some corner
"We found some girl with some guy standing in some corner" (P10262)

In English this is typically translated as an equivalent to 'some (sort of)' and Caubet (1983) concurs with this, adding that it may also refer to size, when used with a singular noun, or quantity when with a plural. This form is invariable and found with singular or plural nouns. All three MCA indefinite structures—bare nouns, waḥəd (l-), ši— appear with French nouns in the current corpus.

Borrowed nouns that are established in MCA generally follow the syntax of MCA described here, but certain borrowings diverge significantly from this system. Words from Berber origin may be definite without any marking, such as with *atay*, 'tea,' and bəllarž, 'stork.' Harell (1962) notes that such words cannot be predicted and must be memorized. Heath (1989) suggests that the existence of these Berber borrowings with zero markings in a definite context establishes a pattern for inherently definite nouns, but notes that this is rare in his data. These noun uses appear to violate the morphosyntax of both languages, rendering them difficult to integrate into a unified syntactic account of the language.

Other unusual borrowing patterns have been observed in MCA. A small class of well-established French borrowings are always used with the French definite article. These borrowings are more often feminine in French such as *lagar*, from 'la gare' is the primary word used to mean 'train station' and its form does not seem to vary (Heath 1989:36, Nortier 1995:92). Finally, in a few cases the leading /l/ of a borrowing has become reanalyzed as the MCA definite article. In this way the Spanish word 'litro' has become /itru/ to denote a single liter (Heath 1989:130). Although these patterns have not been found to be productive, their existence is reason to carefully examine any apparent divergences from expected markings of definiteness with potentially borrowed or code-switching data.

To summarize, there are some similarities between definite determiners in French and MCA; both encode definiteness in a single way, but French employs a single indefinite

form whereas three are possible in MCA. The typical cases of noun feature marking are summarized in Table 7.1.

Table 7.1 Summary of noun feature marking in French and Moroccan Arabic

	French	Moroccan Arabic
Definite	Expressed via article	Expressed via article (on noun and adjective)
Indefinite	Expressed via article	Expressed via article (on noun only - /ʃi/ or /waħəd l-/) OR by lack of article
Gender	Expressed via singular article	Not expressed via noun, expressed in adjective ending
Number	Expressed via article	Expressed via form of noun and adjective

7.3.2 Demonstratives

Demonstratives differ greatly in French and Arabic, and their use in code-switching data has been noted by many in the field (Nortier 1995, Boumans 1998, Mahootian and Santorini 1996). French demonstratives are structurally very similar to articles, seen in 7.7

7.7 a) Cette fille(-ci/là)
The girl

b) Ce garcon(-ci/là)
The boy

c) Ces filles(-ci/là)
The girls

d) Ces garçons(-ci/là)
The boys

Again, demonstrative determiners display gender when singular and number when plural. French demonstratives on their own do not encode information about proximity; this can

be indicated through the addition of '-ci' and 'là' for proximal and distal meanings respectively.

As in MSA, Demonstrative Determiners in MCA select for a full Determiner Phrase, not a noun as in French. This can be seen in 7.8.

7.8 a) had/hada/(ha)dik l-bənt
the=girl

b) had/hadi/(ha)dak d-dərri
the=boy

c) had/hadu/(ha)duk l-bnat
the=girls

d) had/hada/(ha)duk d-drari
the=boys

The Demonstrative form 'had,' described as 'unstressed' by Brustad (2000:115) is fixed and, similar to the Definite Determiner, does not encode number or gender. The proximal forms of the Demonstrative, hada/hadi/hadu, express gender on the singular form and number on the plural as in French. The distal forms, (ha)dak/(ha)dik/(ha)duk, similarly express number or gender. In the case of plural nouns, number continues to be expressed through the vowel pattern in addition to the form of the demonstrative.

The differences between the DEM in French and MCA is substantial; French DEM selects for a noun as a complement, while the MCA DEM takes a Definite Determiner Phrase. This may result in mismatches between the languages, as seen in 7.9

7.9 a) šnu had les masses?
What DEM the masses?
"What are these 'masses'?" (P10261)

b) […] zaʔid hadak l-module dyal S trwa
[…] plus that the-class of S [semester] three
"[…] plus that class from third semester" (P11226)

In 7.9a a full French DP is the complement of the MCA *had* whereas in 7.9b the Arabic definite article is used with the French *module*. The opposite case of a French demonstrative with an Arabic noun (hypothetically, *cette* bənt *this girl) is not attested in the current corpus, or any others that have been identified (cf. Bentahila and Davies 1983, Mahootian and Santorini 1996)

7.3.3 Adjectives

Adjectives represent another point of divergence in the target language pair. The majority of French adjectives are post-nominal, with a small but frequent class that are found pre-nominally, seen in 7.10.

7.10 a) la petite fille intelligente
The small.F girl intelligent.F
"The intelligent small girl"

b) un petit garcon intelligent
a small boy intelligent
"the intelligent small boy"

c) ces petites tables vertes
DEM-PL small-F-PL tables green.F.PL
"those small green tables"

French adjectives must agree in gender and number with the noun that they modify. In the case of plural nouns, as in (7.10c) number and gender can be expressed on the adjective, unlike on the determiner. French adjectives encode only gender and number; seen as feminine singular in (7.10a) represented orthographically as an added final <e>, masculine singular in (7.10b) and feminine plural in (7.10c) as an added final <es>. The adjectives do not show agreement with the definite feature of the noun.

Adjectives in MCA are post-nominal and must agree in number and gender with the noun they modify. In addition, they agree in definiteness, seen in 7.11.

7.11 a) l-bənt d-dkiya
The-girl the-smart.F
"The smart girl"

b) waħəd d-dərri zwin
one the=boy pretty
"A pretty boy"

c) duk ṭ-ṭablat ṣ-ṣɣar l-xḍəṛin
those the-tables l-small.PL the-green.PL
"Those small green tables"

Each of the nouns in (7.11) reflects a different type of definiteness: (7.11a) is definite, (7.11b) is indefinite and (7.11c) has a demonstrative determiner. Assimilation of the definite article /l-/ to an initial coronal consonant is seen in on the adjectives *dkiya* and *ṣɣar* and on the nouns *dərri* and *ṭablat*. This assimilation is aurally salient through a lengthening of the initial consonant. As a contrast, the adjective in (7.11b) contains no the definite marker because the noun that it modifies is indefinite. In (7.11c) *table*, *small* and *green* are all marked for definiteness, but show different allomorphs of /l/. While the French plural adjectives display agreement in both number and gender, in MCA number agreement overrides gender agreement on adjectives. The adjectives *dkiya* and *zwin* both show gender, while the forms of *ṣɣar* and *xḍarin* display plurality but not gender. Those familiar with MSA agreement rules will note that, in MCA, there is no distinction made between human and non-human plurals; all nouns receive the same plural adjective form, the same that is used to mark human male plurals in MSA.

In this way we see that adjectives supply morphosyntactic information about a noun. In both languages, a written adjective form is a further indication of the number feature, although this information is not pronounced on most French adjectives. MCA adjectives display supplemental information regarding definiteness while those in French indicate gender at the same time. If morphosyntax is related to the status of a lexical item

as a borrowing or code-switch, the status of a token will be confirmed through the language of the observed morphosyntax, including that of the adjective. There are few instances in the current corpus where adjectives are present in a CS context, but such cases may help to indicate the level of morphosyntactic integration of French nouns.

Thus, a French noun found with the MCA article /l-/ displays MCA morphosyntax and would expected by Poplack and colleagues (Sankoff and Poplack 1981, Poplack, Sankoff and Miller 1988) to be a borrowing. If an adjective were present on such a noun, it would also be expected to follow MCA morphosyntax by carrying a definite article. Use of a French definite article on a noun with an MCA adjective marked for definiteness in any language would therefore be unexpected, as in the hypothetical example 'la gare l-kbira' *the large train station*, because it would reflect the morphology of both languages.

7.4 FRENCH NOUNS AND NOMINAL STRUCTURES IN THE CURRENT CORPUS

All of the nouns used by 5 or more participants, whether in speech or writing, are seen in Table 7.1 below. The French lexical items are listed by lemma with the frequency, or raw token count, and diffusion, or number of speakers employing the lemma. These numbers reflect the use of the lemma in written and spoken data. The unsimplified ratio given in the final column expresses the frequency and diffusion together for ease of comparison. The inclusion of certain lexical items in this list requires justification. Some could be adjectives, as *informatique, anglais*, *français*, *sixième*, *physique*, and *privé*. Each occurrence of these lexical items was examined to verify its part of speech and any adjectival usages were excluded from the current analysis.

Table 7.2 Most Frequent French nouns in full corpus as lemmas

Lemma	English	Frequency	Diffusion	Freq:Diff
fac(ulté)	campus	43	19	43:19
prof(esseur)	teacher/professor	19	11	19:11
master	Master's degree	21	10	21:10
PC	computer	22	10	22:10
mathématiques	Mathematics	17	9	17:9
an/année	year	24	8	24:8
informatique	computer science	14	8	14:8
salle	room	11	8	11:8
anglais	English	20	7	20:7
français	French	11	7	11:7
message	message	14	7	14:7
sciences	science	8	7	8:7
semestre	semester	23	7	23:7
sixième	sixth	14	7	14:7
économie	economics	14	6	14:6
exam	exam	6	6	6:6
licence	3-year university degree	13	6	13:6
mémoire	thesis	7	6	7:6
physique	physics	11	6	11:6
privé	private	12	6	13:6
ami	friend	7	5	7:5
chimie	chemistry	7	5	7:5
connection	connection	6	5	6:5
conversation	conversation	7	5	7:5
cour	course	13	5	13:5
DEUG	2-year university degree	7	5	7:5
jour	day	9	5	9:5
mention	honors	7	5	7:5
système	system	6	5	6:5
temps	time	7	5	7:5

7.4.1 Illustration of analysis

Detailed analyses of the five most common French nouns are considered first, followed by a summary of the overall results. The most common noun lemma is *fac* (originally from *faculté*), <university>. It is noteworthy that this noun, employed be 19 unique speakers, appears all 43 times as the shortened form *la fac* and always with the French definite article. This includes 4 instances of *had la fac* as a demonstrative DP. The Arabic translation equivalent جامعة /ʒaːmiʕa/ is entirely absent from the corpus, although the adjectival form جامعية is used three times in the context of university residences.

Table 7.3 Syntactic incorporation of '(la) fac'

Mode	Count	French definite	French indefinite	Arabic definite	Null indefinite	Null definite
Spoken	36	36	0	0	0	0
Written	7	7	0	0	0	0

On the surface, it is surprising that a noun that consistently displays French morphosyntax would be the only term used for an institution central to the daily activities of most of the participants. However, only two participants study in fields that are taught primarily in Arabic; the strength of the association between French and university studies may prompt the use of the French term. The use of the French definite article may not be intentional; instead it is possible that *la fac* is a fixed form, better represented as /l̩afak/ and similar to the borrowed form /l̩agaɾ/ 'train station,' which has been lexicalized with the French article as part of the word. Despite the French morphosyntax, the complete lack of an Arabic form for this concept, particularly among university students, suggests that *la fac* is now the preferred term in MCA.

The second most common lemma is <professor> and is used 19 times by 11 speakers in the corpus. All of the uses are the shortened form 'prof,' which in spoken French can be masculine or feminine based on the sex of the individual that it describes.

An interesting note is that two of the spoken uses that display an Arabic definite article also display the feminine Arabic ending /a/ given in 7.2.

7.2)	l-profa	ka-ddir	n-na	*renforcement*	*linguistique*	
	The-professor.F	PRES-she.does	to=us	*reinforcement*	*linguistic*	
	The professor teaches us language training					(P12262)

These are both used by the same speaker, a female engineering student highly proficient in French. Here we see the effect of number; the plural uses of the lemma, and only the plural uses, seem to display a French article, changing from /lə pɾof/ to /le pɾof/. This may be an indication that this word retains a connection to the French language, or it may be that speakers have adopted the vowel as part of an alternative plural formation pattern, which may be separated into the article /l-/ and the plural noun /epɾof/[26]. The MCA translation equivalent استاذ/ة /u:sta:d(a)/ appears 45 times in the data, including 5 times as the plural اساتذة /asatida/. Because this lemma is still realized more often in MCA than in French among the individuals surveyed, its usage in French could conceivably be viewed as an instance of CS in this community. It is also possible that the use of one form over the other carries a distinct social meaning or that the forms might vary in denotation in ways that are still obscure; only two participants use both forms, and they happen to be brother and sister. What we can say for the moment is that despite its high level of morphosyntactic integration, *le prof* is not the only term used in MCA and is not the preferred term for the current speaker group as a whole.

Table 7.4 Syntactic incorporation of '(le/la) prof'

Mode	Count	French definite	French indefinite	Arabic definite	Null indefinite	Null definite	Title
Spoken	13	5 (all plural)	2 (1 plural)	1	3	0	2
Written	6	1 (plural)	0	4	0	1	0

[26] Thanks to Mike Turner for this suggestion.

The third most common French lexical item, PC, <computer>, is masculine in French and provides a useful counterpoint to the feminine *la fac*. Due to the phonetic similarity between the French term /pese/ and the English /pisi/ each spoken token was analyzed auditorily for the presence of have a French-like or English-like vowel and three instances of English-like vowels, all from participants who used English at other points in their speech, were excluded. The information in Table 7.4 gives the morphosyntactic contexts in which the lexeme was found. A stark constrast is found between <computer> and <university>*,* as here, 16 of the 22 tokens of the word appear with an Arabic definite article, including ten of the 15 written tokens. Null indefinite articles, consistent with Arabic syntax, are also used twice in writing. Finally, one null article is found in a context that is clearly definite, violating the syntax of both languages. The Arabic translation equivalent حاسوب /ħa:su:b/ is entirely absent from the corpus. Given its robust integration into the syntax of the Arabic DP and the absence of an Arabic equivalent for this lemma*, PC* appears to be a strong candidate for borrowing in MCA.

Table 7.5 Syntactic incorporation of '(le) pc'

Mode	Total Tokens	French definite	French indefinite	Arabic definite	Null indefinite	Null definite
Spoken	7	0	0	6	0	1
Written	15	0	0	10	3	2

The next most frequent French lemma is *Master*, <master> which is used 21 times by 10 speakers. This lemma is used only as a noun to mean a Master's degree. While the term may have its roots in English, it is used in the French university system and was likely introduced in Morocco from French. Despite the French source of <master>, the morphosyntax is strongly MCA. The MSA equivalent, ماجستير, is absent from the corpus, indicating that this lemma, then, is also likely a borrowing that has displaced the native

lexeme, if one has ever been in use. This would be expected based on the near-categorical use of Arabic morphosyntax with this lemma.

Table 7.6 Syntactic incorporation of '(le) master'

Mode	Total Tokens	French definite	French indefinite	Arabic definite	Null indefinite	Null definite
Spoken	18	0	0	13	4	1
Written	3	0	0	1	0	2

The high use of *an/année*, <year>, as tabulated in Table 7.7, is surprising given that it is not a term, that in its general denotation, is clearly associated with a French-dominated field. However, it appears that they lemma may carry a specialized meaning in that it was used 24 times by 8 speakers, as 'année', but always in the context of school year. The syntactic contexts of this lemma are also surprising because it appears with a high occurrence of null definite articles, which violates the syntax of both languages. The French and Arabic article before a vowel are both realized as /l/. However, the lack of the definite article might be explained by the fact that 15 tokens, marked by an asterix in Table 7.6, occur with an ordinal number, and no article is given for the full noun phrase, seen in (7.3).

(7.3) f *première* *année* dyali
In *first* *year* of=mine
"In my first year [of university]" (P12261)

In this example there can be no doubt that a definite meaning of 'première année' is intended as the MCA genitive particle *dyal* is employed, which must be preceded by a definite noun phrase (Boumans 2006). While this does not explain the syntactic violation, as both languages require a definite article in this context, it suggests that the syntactic violation with this lemma is limited to one specific noun phrase context. In all of these cases, the number refers to a year in university studies, suggesting that année may have taken on the meaning of 'year of university study.' A final fact points to <year> being semantically marked by MCA speakers. The five uses with the null indefinite, all from the

same conversation, are used as ‘année blanche,’ a term that roughly translates to ‘lost year’ to describe a year spent doing studies or other activities that will not contribute to degree completion.

Table 7.7 Syntactic incorporation of ‘(l’)année/an’

Mode	Total Tokens	French definite	French indefinite	Arabic definite	Null indefinite	Null definite	Quant
Spoken	21	0	2	2 * (2 DEM)	5	8 *	4
Written	3					1 *	2

Support for the potential semantic restriction of *année* is strengthened by consideration of the Arabic translation equivalents, which appear more often. Two words for for ‘year’ can be found in Standard Arabic, سنة /sana/ and عام /ʕa:m/, both are used in MCA. /ʕam/ (in MCA) and its dual (/ʕamayn/) and plural (/ʕwam/) forms appear 114 times in the corpus. /səna/ (again, MCA pronunciation) and its MCA plural form /sənin/ appear 21 times. /ʕam/ is used many times to refer to a relative year using ‘next’ or ‘last,’ but just four times with an ordinal number in Arabic, supporting the possibility that *année* is used in a substantially restricted context. Although this French lemma is relatively frequent in the corpus, used by 8 speakers, the Arabic equivalents are used by 28 speakers. Despite its frequency, this word is used with predominantly null definite articles, violating the morphosyntax of French and Arabic. It is impossible to conclude whether the lemma is borrowed with a particular semantic specification, or part of the university student lexicon, a type of university jargon. A comparative speaker group from outside the university or academic context would be needed to answer this question.

7.4.2 Summary of results

The target French-origin nouns are given in Table 7.8 with a summary of the analysis. There are no Arabic translation equivalents for nineteen of the French nouns.

These nouns are likely part of the everyday MCA lexicon of the participants, but are not necessarily fully adopted into MCA. Many of the nouns are used in restricted contexts; several university academic subjects are present including *anglais*, *chimie*, *économie*, *informatique*, *mathématiques*, and *physique*. The lemmas *français*, and *sciences* also fit into this category despite the fact that the Arabic translation equivalents are also used. The two uses of *fransia* in MCA in this corpus are also used in context in which they refer to the French as an academic subject. All of these subjects are taught in French at the university level. These subject names are considered at the least part of a university jargon, but it is impossible to speculate as to whether they have become adopted for these subjects by a wider group of speakers.

Other terms associated with university and secondary studies are also used strictly in French in the corpus. Three refer to degrees that are named as they are in the French system: *DEUG*, *licence*, and *master*. Although they share a semantic domain and each have a French success rate of 100%, these nouns each have a different primary morphology. *DEUG* is generally used in definite contexts without an article (although two instances do have /l-/) while *master* displays the MCA /l-/ on 17 of its 21 tokens. However, the morphosyntax of *licence* is primarily in French (8/13 tokens) but also appears once in speech with a MCA definite article and twice in an indefinite context without an overt article. The final two uses, both in writing, bear no overt article and are in definite contexts, violating the syntax of both languages. By referring to the combined indications of the four metrics, the three degree titles are all considered strong candidates for borrowings despite the non-Arabic morphosyntax of *licence* and *master*.

Five other education-related lexical items have a 100% French success rate: *mémoire*, *mention*, *privé*, *semestre*, and *sixième*. The exception to this semantic grouping is *privé* for which four of the twelve tokens refer to the private sector or business. Although

these concepts appear categorically in French, the morphosyntax within this group again varies widely. One word does not follow the morphosyntax of either language, <thesis>, while three are evenly split between violating morphosyntactic expectations of both language, and following those of Arabic (<private>, <semester> and <sixième>). The final noun, <honors>, follows the morphosyntax of both languages as it is always used in the phrase "with honors" that does not require an article in either language, as in (7.4).

7.4) AKhsna n*validi*w b *mention*
Yes, must=we 1st.*validate*.PL with *honors*
"Yes, we have to pass with honors" (P11292, written)

Despite the high level of variation in their morphosyntax, these nouns are likely borrowed and may have undergone semantic specification as they are only used in educational contexts.

Four French terms are also related to the domain of education but with a lower French success rate: *cour*, *examen*, *prof* and *année*. The French success rate for each of these indicates that the Arabic translation equivalents are used approximately as often, in the cases of *cour* and *examen* while *prof* and *année* are much less common than their Arabic counterparts. Thus the French success rate of these nouns points to a status as code-switches. The unclear morphosyntax of these nouns, equally Arabic and French for *cour* and *prof* while ambiguous for the vowel-initial nouns supports this conclusion. It may be that these French items are in competition with the MCA forms or that use of these terms in one language has implications that are not evident in the corpus; additional data is necessary to know if either of these suppositions is correct.

The other identifiable topic among the French-origin nouns is technology. Four nouns are related to this domain, and particularly refer to online communication: *connection*, *message*, *PC* and *conversation*. While these lemmas are not always used in the

context of technology in monolingual French, they are restricted to that context in the current corpus. *Connection* is used to refer to Internet connection in all of its uses, as are all eight tokens of the related verb *connecter*. The use of *message* is restricted to Facebook messages, and *conversation* is found in a similar context, employed only during the written task to refer to the Facebook conversations in which the speakers were engaged. These words are suspected to be borrowings with semantic specification in the domains of technology and online interaction.

The final five French-origin nouns are difficult to categorize for conversational topic. *Système* is sometimes used in a context of the university system, but this meaning is found for only three of the six tokens while the other refer to systems more generally. When *la salle* is used without any modifier or compound, it references *salle de sport*, or gym, as observed in six of the eleven uses and can be anecdotally heard in this usage in everyday life. The other five tokens of *salle* are used in compound nouns such as *salle d'études*, or in the mixed form *la salle dyal théatre*, 'theater hall,' in which the MCA form *dyal* links the two French nouns. Therefore, it appears that there is a semantic specification for unmodified *la salle* in its sense of 'gym'. The use of the less common French-origin nouns, *jour, temps* and *ami* are understood to be code-switches given the low success rate of the French lemmas relative to their Arabic counterparts.

Table 7.8 French nouns that are potentially borrowed in MCA listed in order of success rate

French Lemma	English	Freq:Diff	MCA translation equivalents	Arabic Frequency and Diffusion	Total lemma tokens	French success rate	Primary Language of Observed Morphosyntax
anglais	English	20:7	انجليزي / انكليزي	0	20	100%	Ambiguous (/l/)
chimie	chemistry	7:5	شيمياء	0	7	100%	Neither
connection	connection	6:5	اتصال	0	6	100%	Arabic/Neither
DEUG	2-year university degree	7:5		0	7	100%	Neither
économie	economics	14:6	اقتصاد	0	14	100%	Ambiguous (/l/)
fac(ulté)	campus	43:19	جامعة	0	43	100%	French
informatique	computer science	14:8	معلوميات	0	14	100%	Ambiguous (/l/)
licence	3-year university degree	13:6		0	13	100%	French (some variation)
master	Master's degree	21:10	ماجستير	0	21	100%	Arabic
mathématiques	Mathematics	17:9	رياضيات	0	13	100%	Arabic
mémoire	thesis	7:6	رسالة	0	7	100%	Neither
mention	honors	7:5		0	7	100%	Ambiguous
message	message	14:7	رسالة	0	14	100%	Arabic
PC	computer	22:10	حاسوب	0	22	100%	Arabic
physique	physics	11:6	فيزياء	0	11	100%	Arabic/neither
privé	private	12:6	خاص	0	12	100%	Neither/Arabic
semestre	semester	23:7	فصل	0	23	100%	Neither/Arabic
sixième	sixth year of secondary school	14:7	سادس	0	14	100%	Neither
système	system	6:5	نظام	0	6	100%	Arabic

Table 7.8 (cont.) French nouns that are potentially borrowed in MCA listed in order of success rate

French Lemma	English	Freq:Diff	MCA translation equivalents	Arabic Frequency and Diffusion	Total lemma tokens	French success rate	Primary Language of Observed Morphosyntax
français	French	11:7	فرانسية	2:2	13	84.60%	Arabic
sciences	science	8:7	علوم	2:2	10	80%	Neither
salle	room	11:8	غرفة	0	14	78.60%	French
			بيت	3:4			
cour	course	19:10	صف	0	31	51%	Arabic/French
			قسم	12:4			
conversation	conversation	7:5	حوار	0	14	50%	Neither
			هدرة	7:6			
exam	exam	10:8	امتحان	9:7	19	53%	Ambiguous (Plural=French)
jour	day	10:6	يوم	14:7	24	42%	French
prof(esseur)	teacher/professor	19:11	استاذ	32:14	51	37%	French/Arabic (equal)
an/année	year	24:8	سنة	21:8	159	15%	Ambiguous
			عام	114:30			
temps	time	7:5	وقت	40:15	47	15%	French
ami	friend	7:5	صاحب	116:26	123	6%	French

7.5 DISCUSSION

The results of this investigation, examining the 30 most diffused French nouns in the corpus, are summarized in Table 7.7. Each French lemma is given with its logical MCA translation equivalents. The frequency and diffusion of the lemmas is shown as well, in addition to the use of the MCA translation equivalents and finally the French success rate. It can be seen in Table 7.7 that no single measure included in this analysis can be used to determine whether a French-origin lexical item is likely to have become a part of Moroccan Arabic. The first metric, frequency, may be misleading in situations such as *an/année* in which the French term is used more often than other French terms, but much less often than the Arabic equivalent. The second metric, diffusion, reveals that only the most common French-origin item, *la fac*, is used by more than half of the speakers (19/36). The French lexical item with the second highest diffusion rate is *le prof*, for which there is not strong evidence of borrowing and is used by 11 speakers, a much lower number than *la fac.* In addition, the concept of <professor> seems to be associated with a single language for most speakers; only two participants use the term in both languages, while the other 23 use it only in one language.

The third measure, observed morphosyntax, is sometimes used to distinguish borrowings from CS at the token level (Poplack and Meechan 1988, Poplack et al. 1988). On its own, this metric may misrepresent borrowing status as morphosyntax is complicated in this language pairing by the surface similarity of the definite articles. Even when the definite article is clearly French, it does not indicate the word's status. The most frequent French-origin item, *la fac*, is used exclusively with a French definite determiner but is the only term used for this referent among a group of primarily university students. The definite

article also cannot morphologically distinguish nouns that begin with a vowel, rendering a distinction between French and Arabic definite articles impossible in this context. The same is true of null definite articles; these are common for many of the identified French lemmas and are the most common morphosyntax for six of the lemmas analyzed. Yet it is impossible to know whether these lemmas have joined the small set of inherently definite nouns in MCA that do not take morphological marking or whether speakers may simply avoid article usage, as has been suggested in other contact situations (Otheguy & Lapidus 2003). For six of the 30 nouns analyzed, there is a high level of variation in the morphosyntax of the noun, which may denote that the status of the lemma is undergoing change. Conversely, variations in morphosyntax could also be indicative of differences between speaker's preferences and be irrelevant to the status as a French item as a borrowing or code-switch.

When considered across the French nouns investigated here, it is clear that morphosyntax is not a reliable indicator of borrowing in this language pairing. If this were to be the sole indicator of borrowing, all uses of <university> would be analyzed as code-switches as they all display a French article despite the fact that only the French term is used in the corpus. This metric alone would also split the uses of <professor> along the number feature as the singular uses of the term appear with an Arabic article, and would thus be 'nonce borrowings' using Poplack and associates' terminology (Poplack et al. 1988, Poplack & Meechan 1998) while the plural uses all display French morphosyntax through the use of the French article *les*. These observations of the morphosyntax in the corpus under study are a strong argument against the use of this metric at the level of token for this language pairing.

The final measure, presence of MCA translation equivalents, is perhaps the strongest indicator of borrowing in the corpus. When the lemma itself is considered and its overt manifestation in each language is compared then the success rate of the French-origin lexical items can be determined. However, clearly we need more data since the overall rate of borrowing is quite low and some French origin tokens are sparsely represented in the corpus. A much larger corpus with a wider variety of participants and participant backgrounds is necessary in order to truly understand the current make-up of the Moroccan Arabic lexicon.

These results of the current analysis suggest that lack of translation equivalent is the best indicator of the status of French-origin nouns in MCA. However, no one factor on its own is sufficient to determine this. Such a finding recalls Mackey's (1970) notion of 'availability' as an indicator of noun borrowing. Testing the range of a bilingual individual's lexicon in a more directed recall would be one way to assess the language in which certain lemmas are most readily accessed by bilinguals. This type of investigation with a larger cohort set of participants might indicate whether the nouns investigated here are used by a broader sample of speakers in Morocco or if they have the status of jargon, used strictly by young adults of the current sample, and more specifically by current and recent university students.

The results in this chapter again emphasize the overall dominance of Moroccan Arabic in the corpus. Chapter 6 demonstrates that the speakers use very little French in their production and this analysis further suggests that close to 300 of the French lexical items originally categorized as instances of CS may instead be better viewed as part of the MCA lexicon. If these lexical items were included as borrowings in the previous analysis,

this would represent a small decrease in the rate of French by all speakers as well as a decrease in the number of French-origin nouns and DPs noted in the corpus. Work in identifying borrowings into MCA is significant as a way to provide an accurate picture of the use of MCA-French CS used by native speakers of MCA. These borrowings remain of French origin, but if the lexical items are accessible to monolingual native speakers, and part of MCA, they are evidence of the historical influence of French on this dialect and not of its current prominence in the country. The level of French-origin lexical items present in MCA could then be compared with those of other Arabic dialects, whether of French, English, Italian or other origin, to determine whether the level of influence from French is quantitatively greater in MCA or whether this characterization is an unfounded stereotype of the western North African dialects.

8. Discussion and Conclusion

The primary aim of this dissertation is to quantify the use of French in MCA-French code-switching by young adults. This is examined through the rate, or proportion of French lexical items used, in addition to the syntactic constituents in which French lexical items are found. The results indicate throughout the analysis that the majority of participants use very little French; 25 of the 34 participants use a rate of less than 4.5% French lexical items in speaking and while the written French rate is higher for almost all participants, it remains below 10% for 21 of the 28 participants. Right away, rate of French indicates important differences between language modalities for the participants. The consideration of constituent switched reveals other differences by modality; a greater proportion of Nouns/Noun Phrases and Determiner Phrases are used in the spoken data, while extra-sentential elements such as 'oui' and 'non' and interjections are more common in the written IM data.

In addition to differences by modality, the impact of the extra-linguistic factors of SEX, FRENCH PROFICIENCY and LANGUAGE ATTITUDE were also considered. These factors are seen to have an impact on MCA-French CS as described in Chapter 6 and summarized in §8.1 below. The inclusion of both rate and constituent switched reveals differences in how the selected factors impact use of CS, as certain differences were found in constituents switched and others in rate of French lexical items.

The current findings begin to answer the target research questions, and carry broader implications for the fields of language contact and CS. The results presented here raise numerous lines of future inquiry in order to better understand Arabic-French CS in

this language pair in addition to the phenomenon of CS across language pairings and communities.

8.1 RESEARCH QUESTIONS AND THEORETICAL CONTRIBUTIONS

The experimental tasks address the research questions formulated in Chapter 5 above and repeated here with our primary findings.

Question 1. What proportion of informal language use by young Moroccans is constituted by French lexical items?

Summary of results: The rate of French lexical items in MCA-French CS varies widely in the data set and is best viewed by modality, as addressed in Question 2 below. However, the overall finding is that young adult speakers of MCA use very little French, in contrast to the common stereotypes that portray Moroccan Arabic as containing a high amount of French.

Question 1b. What is the structure of the French lexical items in MCA?

Summary of results: Bare nouns and noun phrases are by far the most common French constituent type throughout the corpus. Nouns and noun phrases are employed by almost all speakers in both modalities and in the spoken mode represent close to one half of all French constituents used in the spoken data. Many of these nouns are accompanied by an MCA article, while others carry no overt article. The occurrence of other constituent types differs notably between modalities and for this reason is discussed in greater detail in the discussion of Question 2 below.

Question 2. Does communication modality (written or spoken) affect the rate or structure of French in MCA-French CS?

Summary of results: There is a vast difference between code-switching practices in spoken conversation and in written IM. The use of French in speech is very low among most participants, with a 2.6% median rate of French. The spoken data shows an overall lower rate of French use, but there is a broader range in use between speakers, from a low rate of 0.5% French lexical items, up to 33.99%. The small group of high French users raise the mean rate to 5.6%. The rate of French in written IM shows a lower degree of variation; the median rate is 8.2% and the mean is only slightly higher at 9.2%, but with an overall range of 1.2% to 30.0% French. The similarity of the median and mean in the written data indicates that there are fewer outliers in the written corpus.

The structure of MCA-French CS is also affected by modality. While the category of Noun/Noun Phrase is the most common constituent type across modalities, this category accounts for a substantially different proportion of French constituents in each: 32.2% of the IM data and 45.9% of the spoken conversation. The primary French constituent type in the spoken corpus is nouns, either without a French determiner, as in the constituent type Noun/Noun Phrase, or with one in a full French Determiner Phrase; DPs constitute an additional 29.8% of the spoken corpus. This contrasts sharply with the IM data, in which only 9.0% of the French constituents are Determiner phrases, making this category the third most common constituent type in written IM. The second largest constituent type in the IM data is instead the affirmative/negative particles 'oui'/'non,' which are almost 9 times as common than in the spoken conversations. Other constituent types make up much smaller proportions of the French data in both modalities.

Question 3. How do external factors of SEX, FRENCH PROFICIENCY or LANGUAGE ATTITUDES contribute to the variation observed in rates and structure of CS?[27]

Summary of results: The rates of French lexical items show distinct trends by extra-linguistic factor and modality. A significant difference is found between the rate of French used by male and female participants in written IM, but not in speech. Surprisingly, it is the male speakers who use more French in written IM, despite the fact that stereotypes predict that females would be higher French users. Another surprise is found in the comparison of rate of French and proficiency. There is no relationship between rate of French and French proficiency in either modality, even when two measures of French proficiency (pen-and-paper grammar test and self-rating) are included. This conflicts somewhat with the findings of Sayahi (2011) whose proxy measure of French proficiency, level of education, does indicate an association between this factor and CS. It is possible that proficiency has an effect among his speakers, but not the current sample; it may also be that it is level of education, and not French proficiency, that impacts the use of CS in Tunisia. Finally, LANGUAGE ATTITUDE also displays divergent results by modality. There is no impact of LANGUAGE ATTITUDE on written production, which may indicate that CS in writing is part of the written conventions, while there is a significant impact of this factor in spoken conversation, in which those with the most positive attitude toward French use the highest proportion of French lexical items. All of these results point to the importance of including quantification of rate of CS and extra linguistic factors in a complete analysis of this phenomenon in any language pair.

[27] These terms are given in small caps when used as extra-linguistic factors and in standard type face when used to refer to characteristics of a given speaker or speakers.

The hypotheses with regard to constituent type and extra-linguistic factors of SEX, FRENCH PROFICIENCY and LANGUAGE ATTITUDES were tentative because of the dearth of past studies on constituents switched. It was predicted that SEX would affect constituents switched, but exact expectations were not given due to the existence of only a single study (Sayahi 2011a) on the topic. We expected FRENCH PROFICIENCY to be related to a greater diversity of French constituents, but no hypotheses could be formed regarding LANGUAGE ATTITUDE and constituents switched. A statistical analysis of constituents switched in French was not possible because of the small numbers, but visual inspection of the data based on proportion of constituents used by speakers implies differences by extra-linguistic factor. Furthermore, analysis of extra-linguistic factors by constituent type reveals distinctions between groups that were not visible in the analysis of rate alone. For this reason, a quantitative analysis of constituent type should be included in future studies of CS in order to determine whether the trends noted here are specific to the current corpus or can be found across language pairs.

SEX is found to impact the French constituents used, but in the opposite direction of past findings. Males employ a higher proportion of Nouns and Noun Phrases than females in both writing and speaking. The male speakers in Sayahi's (2011a) data set use a lower rate of this constituent type than the female speakers. The divergence between these two data sets suggests that use of French is impacted by community norms. In the current corpus, another difference exists in the use of Oui/Non and Simple Phrases, both of which are used more often by female speakers.

As expected, FRENCH PROFICIENCY also impacts the constituents used: higher proficiency is related to greater constituent variety across modes. In the IM data, the use of

Determiner Phrases increases with proficiency in French while the use of Simple Phrases and Interjections is highest among speakers with lower proficiency levels. There appears to be a stronger impact of FRENCH PROFICIENCY on how French constituents are used in spoken language; the high proficiency speaker group uses a greater variety of French constituent types, which mirrors the results found between generations by Bentahila and Davies (1995), and between educational levels by Sayahi (2011a). Each of these studies finds that those with more education in French, and thus presumably greater proficiency in it, use a greater variety of French constituents. The French constituents of the low proficiency group show a vast majority of nominal structures (87.8%) while this category represents a substantially lower proportion of the speech of high proficiency group (71.2%). There are also distinctions within the category of nominal structure as the proportion of Nouns/Noun Phrases decreases with a rise in French proficiency while the proportion of Determiner Phrases increases with increased proficiency in French. In short, those with a higher level of French are more likely to use French lexical items with a determiner from the donor language.

The use of Nouns/Noun Phrases is also notable when participants are grouped according to their attitude toward French. Speakers with the most positive attitudes toward French use a lower proportion of Nouns/Noun Phrases in both writing and speech. This is similar to the effect of FRENCH PROFICIENCY, but it is clear that LANGUAGE ATTITUDE as a factor measures something separate from proficiency because the use of Determiner Phrases is not affected by different levels of speaker attitude. Finally, speakers with more positive attitudes toward French use a higher rate of Complementizer Phrases in each mode than those with less positive attitudes; they also use a wider variety of constituent types.

8.2 IMPLICATIONS

The findings of this dissertation regarding the differences between written and spoken CS provide additional empirical support to the calls for a model of CS specific to written language use (Sebba 2012, Kytölä 2012, Lexander 2012). Stark differences are visible in the level of CS found in each modality in addition to the structures in which CS is employed. The analysis of rate of CS by the target extra-linguistic factors further underscores the need to treat written CS separately from spoken, due to the fact that both SEX and LANGUAGE ATTITUDES reveal statistically significant differences in only one modality. The need for a writing-specific model is also noted at a syntactic level; most of the 30 French-origin nouns analyzed in Chapter 7 display the same morphosyntactic environments in written and spoken use, but many do not. Faced with a forced decision regarding article choice, speakers may choose to avoid the issue by omitting one completely. Avoidance of article choice through omission has been found as a strategy in CS between English and Spanish (Otheguy and Lapidus 2003).

The corpus collected for this study presents an opportunity to observe a phenomenon that is long past for most language varieties: the emergence of literacy in a language. While there is a strong literary tradition in SA and in French, the frequent use of Moroccan Arabic in writing, and thus in reading, by a large portion of the population is still in its early stages. The lack of a written standard for this variety may be a boon to the study of CS in this context as there are few expectations regarding orthography or written grammar that speakers may otherwise feel compelled to follow. The use of Romanized Arabic may diminish any expectations that could otherwise carry over from SA, the highly

standardized and codified variety of Arabic associated with Arabic script.[28] The spontaneous style of CS in an IM context also distinguishes it from traditional written CS, whether in literature or personal letters, as planning and editing are minimal in informal chat. It is impossible to predict whether the written form will ever move from Facebook chats and forum comments to print newspapers or books. Egyptian Arabic has a higher presence in writing than other dialects, seen in its use in literature and listing as a language of Wikipedia articles distinct from Modern Standard Arabic, but Egyptian Arabic is written primarily in Arabic script. The common use of Romanized Arabic may prevent Moroccan Arabic from making a transition to other domains.

The current corpus may also present valuable data on incipient loanword integration of some of the many lexical items related to technology. Loan word adaptation is directly linked with technology because it is this referential domain that is supplying the lexical items that are being swiftly adopted, such as the verbs <like>, <share>, and <comment>. These actions are all specific to the domain of Facebook and are relatively rare in the corpus, but the examples identified tend to use a light verb strategy, as seen in (8.1)

(8.1) dir *j'aime* *partager* dir *commentaire* hadi hiya l-bllia
do *I=like* *to.share* do *comment* that she.is the=addiction
"Like, share, comment – that's what the addiction is" (P11214)

Use of a light verb strategy for syntactic incorporation of French verbs is previously unattested in Moroccan Arabic, despite its occurrence with Dutch verbs, and is notable for this reason (Ziamari 2008:229).

The quantification of CS rate allows for quick identification of the overall use of CS within a corpus as a reflection of how widespread the practice is within a speaker group,

[28] However, this could also be tested by comparing RMCA with MCA in Arabic script.

or in the speech of a single individual. This insistence on quantification is becoming common across various sub-fields of linguistics that examine variation in language use. Quantification in CS research will provide a means for simple comparison between speaker groups and between language pairings. The lack of existing quantified analyses of CS may have arisen because of the past difficulties in completing such an analysis, but modern Natural Language Processing methods facilitate the task tremendously.

The current findings will also inform any future work on syntactic theory or a typology of CS as a complete theory must also account for the structures and variation observed in CS. Data from Arabic-French CS is invoked in many discussions of CS due to the typological distance between the two varieties, combined with the extended history of contact. The present data provides a quantitative empirical basis for making and supporting any such theoretical claims. While a high degree of variation is seen in the CS of young Moroccan speakers, very few of the structural observations noted in the preceding chapters can be said to be single examples that may safely be discarded. Future data sources may show that even the infrequent structures are common among a certain subgroup of speakers and cannot be ignored.

8.3 LIMITATIONS

The greatest limitation to the current study is the homogeneity of the sample of speakers analyzed. As noted in the title, this study focuses on university-aged speakers and the results may not be applicable to individuals outside of this limited group. An analysis of a wider sample from the population of Morocco is needed to indicate whether the rates and patterns of CS observed here are typical of speech across the lifespan or are particular

to individuals of the target age range. The sample of fields of study is also unbalanced in the target speaker group. It was noted that the students in Engineering use French in a substantially different way from those in other fields of study, but the low number of Engineering majors precluded these individuals from inclusion in the analysis of extra-linguistic factors. In addition, the most common field of study among participants is English, while many fields of study were completely unrepresented. Two potential solutions exist for this issue; first, inclusion of students from a single field of study would control for differences that may result from area of studies. The second possibility would be to use a broad sample of speakers across a variety of fields of study. The latter option is recommended if the goal is to obtain a sample that is representative of university students more generally.

Region is also limited in this study, as speakers were all recorded in Meknes. Several regional dialects exist in Morocco, with varying level of contact with French, Spanish, and Berber languages (Heath 1989, Aguadé 2010). While it is often found that the dominant dialect of Morocco is that found in urban centers, Casablanca most specifically, dialectal differences remain between major cities (Hachimi 2007). The speakers included in this analysis are originally from a variety of cities and regions, increasing the likelihood that the results of the study are generalizable in the target age group, but an analysis of CS practices in other large cities would confirm this. The contact situation is also somewhat different in other large cities; Casablanca, the largest city in Morocco, is home to many multi-national companies, while the Rabat-Salé metropolitan area has a notable ex-pat population from a variety of nations because of the embassies and consulates found in the capital. Both of these areas are also important tourist destinations,

as are Fes, Marrakech and Tangier, three other major population centers. For these reasons, Moroccan residents of these cities are more exposed to foreign languages and may use French in different ways from the rates and patterns found in Meknes.

8.4 FUTURE RESEARCH

The corpus created in the course of this research provides sufficient data to carry out a lifetime of future projects and raises questions that can be analyzed in other existing data sets or future corpora. It is possible for many of these questions to be explored by specialists in a variety of fields through sharing the corpus created for this dissertation and through the methodologies employed here. Future research possibilities exist in the fields of Arabic dialectology, lexical borrowing, grammatical theories of CS, writing systems, new literacy studies, and phonology, among others. Sharing the full corpus for use and review by other researchers will open this study to critiques and criticism; such criticism will fuel open dialogue in studies of Arabic-French CS and in the broader field.

As noted in the introduction, this research and the related corpus represent a rich source for data on Moroccan Arabic and in this way add to Arabic dialectology. One major project to be undertaken to maximize the impact of this corpus is to extend the transcription to the full 18 hours of recorded conversation in order to identify whether the rates of code-switching attested in the current study are constant throughout the conversations. Each 15 minute segment included in the current corpus was chosen for the inclusion of the topic of education, which may lead to a higher use of French lexical items than might be found in other topics. This decision is also visible in the 30 most common French nouns, noted in Chapter 7, as 20 of these lexical items belong to the semantic domain of education. This

corpus could also be examined to evaluate claims regarding the structure of MCA, such as Turner's (2013) proposal that the apparent definite article in MCA does not, in fact, denote a definite meaning.

The issue of undocumented borrowings, treated in Chapter 7, should be further explored in order to identify whether the lexical items identified as potential borrowings have been fully integrated into the MCA lexicon. While the majority of the 30 French nouns employed by participants are related to the domain of education, it is unclear whether these terms are preferred by a broader community of MCA speakers. If these are the primary terms used by speakers of different backgrounds, including those of other generations and education levels, they are more likely to be fully borrowed into the dialect; if their use is restricted, they may instead be a type of jargon restricted to current university speakers. One way to resolve this question would be to record speakers from a wide variety of backgrounds in directed interviews about educational topics and observe the terms that they use spontaneously. Another possibility would be to verbally elicit word lists related to a variety of topics, including education. For both tasks, the data obtained would be paired with metalinguistic speaker information including age, level of education, type of education completed (private or public), region of education and current residence, and current field of employment. The education-related lexicon derived from either of these methodologies could then be analyzed across speaker groups to identify whether any variation is found for the same concepts between members of different generations, levels of education, education types, or regions.

The variation in French constituent types as used by different speakers exposes a potential challenge to theoretical accounts of CS structure. If CS structure is based strictly

on language-internal factors (that is, on grammar alone), it is difficult to explain the structural differences noted here between speakers of different sexes or speakers who harbor different attitudes.[29] Thus, if there are differences between the sexes in the internal structure of noun phrases containing the same target nouns, it is evidence that such structure is utilized because of social factors and not purely grammatical considerations. An investigation of grammatical structure and social factors would differ from the current analysis as it would necessarily focus more on the structure of select lexical items and whether that structure varies across speaker groups. Such an analysis would require a much larger corpus in order to uncover meaningful statistical differences.

Future study should also focus on the emerging writing practices in Moroccan Arabic. While the current study considers how French is used in the written modality, there is no existing description of the system of how the Romanized alphabet is used to depict Arabic for this dialect. Moroccan is included in Elhija's (2014) cross-dialectal description of written Arabic dialects, but her analysis includes limited data from Moroccan Arabic and the written trends given in that article do not match the tendencies seen in the written portion of the corpus collected for this dissertation. A first step at describing the sound-symbol correspondences is found in §4.3.4, but was not completed systematically. For example, it is clear through perusing Facebook pages that Arabic script is used by many speakers of MCA to represent their dialect in writing, despite its absence from the current study. It may be that each mode of CMC, such as discussion forums instead of IM, is associated with different script use or with a distinct group of speakers who have their own

[29] Proficiency could reasonably affect CS structure as speakers with lower proficiency in one language may not be sensitive to the grammatical rules of that language in general and may therefore produce CS structures that would be absent from the speech of individuals with higher proficiency, as argued originally by Poplack (1980).

unique writing conventions. A greater understanding of differences in CMC modes entails a description of written Moroccan Arabic, including the orthographic system and the written representation of word boundaries. The syntax of the dialect in writing should also be described in order to verify that morphological differences, such as consonant germination, are preserved in this modality. The emerging literacy practices of this dialect also merit further study. Past trends of writing in MCA have never spread to a wide audience, but the communicative contexts afforded by CMC already facilitate wide use of the written dialect among native speakers. In CMC, more speakers than ever are reading and writing their native variety despite a lack of standardized form.

The morpho-syntax of the 30 most common French-origin nouns is detailed in Chapter 7, but further analysis could be completed with the full set of nouns and the morpho-syntactic contexts in which they appear. A variationist analysis of the morpho-syntax of French nouns would indicate whether the observed syntactic variation correlates with the target extra-linguistic factors. Variation in CS, similar to variation in any monolingual context, is likely meaningful. Two variables are implicated in the current data set: sex and language attitudes. The socioindexical meanings of CS can be explored based on the current findings. Is high use of French in IM viewed as masculine? Does greater use of 'oui' and 'non' index femininity? A positive attitude toward French co-occurs with a higher rate of CS in speech, but how is the difference in rate perceived by other speakers? These and other issues may cause us to readjust our understanding of variation in CS and lead us to refine our understanding of CS typologies such as that put forth by Muysken (2000).

Pronunciation of French lexical items may be used to comment on the use of CS. Old stereotypes hold that females use a higher rate of French, but this was not found to be the case in either modality of the current study. Ziamari (2008) notes that the pronunciation of French-origin lexical items by female speakers is more faithful to the source language than that of males. The quality of many of the current recordings will allow for a phonetic analysis of French lexical items in order to determine whether this anecdotal report proves true in the current corpus. Once instrumental observations have been made regarding the pronunciation of French-origin lexical items, future work can explore what variation in pronunciation may index among speakers.

A diachronic analysis of MCA-French CS could be carried out if another data set were analyzed in a way similar to the current data. For example, Ziamari's (2008) data was collected among engineering students in Meknes, Morocco approximately 13 years before the current data was collected. The same location and age group provide a certain similarity between data sets that would facilitate comparison of the two corpora, particularly between her corpus and the spoken production of the four students in Engineering included in the current data set. Evaluation of the corpora together would indicate how CS practices have evolved, or remained constant, among this speaker group.

These topics and many others can be addressed in order to improve our understanding of CS in this and other language pairs. Many varied communities participate in CS in this language pair; focusing on French in contact with MCA in France or Quebec, where Arabic is a minority language, will likely display very different trends due to community language dominance and may also show an effect of proficiency at the level of the individual. Other dialects in contact with French are likely to differ from the current

results, whether Arabic is the minority or majority language of the community, due to differences in use of French in mandatory education, history of colonization, and prevailing language attitudes.

8.5 CONCLUDING REMARKS

A corpus-based approach to Moroccan Arabic-French CS reveals that there are actually few French lexical items used by young adult native speakers of MCA, particularly in speech. In contrast to dialect stereotypes, young adults speak primarily in Moroccan Arabic and restrict their use of French to lexical items that are most often related to concepts and domains in which French is dominant: education and technology. One conclusion that might be tentatively drawn is that the future of French in Morocco is uncertain. This study demonstrates that the use of French is already quite limited in daily conversations for young adults, even among highly educated individuals. Nonetheless, a small number of speakers in this study use a very high rate of French. It may be that many others use a similarly high rate of French but were not included due to the location of the study or the recruiting techniques employed. Future investigations will indicate whether this is the case.

This dissertation set out to apply a quantitative analysis to the code-switching practices of young adults in Morocco in an effort to contribute to our understanding of CS among this speaker group and the field more generally. Quantification of both the rate and structure of CS is necessary to understand how these speakers use their linguistic repertoire and to examine how social and affective factors contribute to variation in CS usage. Here we have learned that use of CS varies greatly by modality. In spoken production, the rate

of CS in this community is, in general, partially dependent on the language attitudes of the speakers, while variations in how CS is used among these same speakers is conditioned instead by their sex and French language proficiency. We hope that future accounts of CS will adapt similar methods to those used here in order to make cross-linguistic and cross-community comparisons of CS possible, including comparisons of the social factors that impact it. We also hope that theoretical work on CS will take into account the richness of this phenomenon.

None of the conclusions presented in this work should be considered absolute. Moroccans will continue to speak MCA, attitudes toward French and the dialect will change, and the written system will continue to evolve through use or fade through lack of use. The current methodology and presentation of results is meant to emphasize the value of quantification of CS and exemplify how automatic processing can be used for language data even when a variety is not well described. Continued comparison of this data with future corpora will help to contribute to the changing nature of CS in this community.

Appendix A: Background questionnaire

Titre de l'étude: La communication entre les jeunes marocains

Ce sondage contient des questions simples sur vous, votre formation, et les langues que vous et votre famille parlez. Vous n'indiquerez jamais ni votre nom, ni votre prénom, ni d'autres informations qui pourraient vous identifier. Des réponses honnêtes et détaillées seront souhaitées et bienvenues.

عنوان البحث: التواصل بين شباب المغرب

في هذا الإستطلاع توجد اسئلة عن حياتك ودراستك واللغات التي تتكلمها عائلتك. لن تكتب اسمك او كنيتك او معلومات أخرى لها رابط معك شخصيا. إجاباتكم الواقعية و المُفصّلة سيكون مُرحّبا بها.

Numéro d'étude: _____ رقم البحث:

Vous : **انت:**

Vous êtes: M ذكر / F أنثى :انت

Date de naissance: _____ تاريخ ميلادك:

Si vous avez vécu dans d'autres villes, veuillez indiquer les deux dernières villes dans lesquelles vous avez habité et pendant combien de temps:

إذا سكنت في مدن أخرى اكتب(ي) اسماء المدينتين الأخيرين والمدة التي قضيتها في كل مدينة:

1: _____ :1

2: _____ :2

Votre formation et les langues				**دراستك واللغات**
Votre école primaire était :	**Publique** / **عام**	**خاص** بالعربية	**Privée** En français	مدرستك أو مدارسك الابتدائة:
Votre (vos) collège(s) était :	**Publique** / **عام**	**خاص** بالعربية	**Privée** En français	مدرستك أو مدارسك الاعدادية:
Votre (vos) lycée(s) était :	**Publique** / **عام**	**خاص** بالعربية	**Privée** En français	مدرستك أو مدارسك الثانوية:
Quel est le diplôme le plus élevé que vous **avez obtenu** ?				ما هي الشهادة العالى التي كاملتها؟

Université ou emploi actuelle:	اسم الجامعة أو الوظيفة الحالية:
Spécialité actuelle:	تخصصك الحالي :
Année d'études actuelle (première de licence, etc.):	السنة الحالي في الجامعة (اول في الإجازة...)
Quel est le diplôme le plus élevé que vous **comptez obtenir** ?	ما هي الشهادة العالى التي تفكر في ان تكاملها؟
Quelle poste comptez-vous avoir en cinq ans ?	أي وظيفة تريدها بعد 5 سنوات؟
En quelle année avez-vous commencé à apprendre l'arabe standard?	في اي سنة بدات تعلم اللغة العربية الفصحى ؟
En quelle année avez-vous commencé à apprendre le français?	في اي سنة بدات تعلم اللغة الفرنسية؟

Votre famille	**عائلتك**
De quelle ville vient votre père?	أبوك من أي مدينة؟
Quelle(s) langue(s) est-ce que votre père parle?	أي لغة أو لغات يتكلم أبوك؟
Qu'est-ce que votre père fait comme travail?	ماذا يشتغل أبوك؟
De quelle ville vient votre mère?	أمك من أي مدينة؟
Quelle(s) langue(s) est-ce que votre mère parle?	أي لغة أو لغات تتكلم أمك؟
Est-ce que votre mère travaille? Si oui, qu'est-ce que votre mère fait comme travail?	هل امك تشتغل؟ إذا قلت نعم, فماذا تشتغل؟

Est-ce que vous avez de la famille dans des pays auxquels le français est langue officielle ? (La France, le Belgique, le Canada, la Suisse, etc)	هل عندك عائلة في بلاد فيها الفرنسية لغة رسمية؟ (فرنسا, بلجيكا, كانادا, سويسرا...)
Si oui, vous leur parlez….	إذا قلت نعم,هل تتكلم(ي) معهم...

كل يوم	كل اسبوع	كل شهر	3-9 مرات في السنة	أقل من 3 مرات في السنة	ابدا
tous les jours	toutes les semaines	tous les mois	3-9 fois par an	moins souvent que 3 fois par an	jamais

Si oui, en quelle(s) langue(s) est-ce que vous leur parlez?	إذا قلت نعم, باي لغة أو لغات تتكلم معهم؟
Avec qui ? ما من؟	Langue(s) لغة

Les langues dans votre maison	اللغات في البيت
Directions: Indiquez la fréquence à laquelle votre famille utilise les langues données **dans votre maison.**	تبين كيف تستخدم أسرتك اللغات **في بيتك**.

	Jamais	أبدا	Rarement	نادرا	Parfois	أحيانا	Souvent	عادة	Très souvent	غالبا	Toujours	دائما	
Mère													أم
Le dialecte marocain													دارجة مغربية
L'arabe standard													عربية فصحى
Le français													فرنسية

	La langue amazigh							أمازيغية
	Mélange dialecte et français							خليط الدارجة والفرنسية
	autre							أخرى
Père								أب
	Le dialecte marocain							دارجة مغربية
	L'arabe standard							عربية فصحى
	Le français							فرنسية
	La langue amazigh							أمازيغية
	Mélange dialecte et français							خليط الدارجة والفرنسية
	autre							أخرى
Frère								أخ
	Le dialecte marocain							دارجة مغربية
	L'arabe standard							عربية فصحى
	Le français							فرنسية
	La langue amazigh							أمازيغية
	Mélange dialecte et français							خليط الدارجة والفرنسية
	autre							أخرى
Soeur								أخت
	Le dialecte marocain							دارجة مغربية
	L'arabe standard							عربية فصحى

	Le français							فرنسية	
	La langue amazigh							أمازيغية	
	Mélange dialecte et français							خليط الدارجة والفرنسية	
	autre							أخرى	

L'internet et les medias	الانترنت ووسائل الاعلام
Quel est votre site web préféré?	ما هو موقع الانترنت الذي تفضل؟
En quelle langue est-ce que vous utilisez Facebook?	في أي لغة أو لغات تستخدم فايسبوك؟
Est-ce que vous tchattez souvent sur internet?	هل تدردش في الانترنت عادة؟
Quelle(s) langue(s) vous utilisez en générale pour tchatter sur internet?	أي لغة او لغات تستخدم في الدردشة في الانترنت غالبا؟
Quel(s) site(s) ou logiciel(s) utilisez-vous pour tchatter?	اي برامج أو موقع تستخدم في الدردشة في الانترنت؟
Est-ce que vous lisez l'arabe marocain écrit en lettres "français/anglais" dans les tchats?	هل تقرأ الدارجة المغربية بحروف "فرنسية/انجليزية" في الدردشة ؟
Si oui, qu'est-ce que vous pensez des gens qui écrivent comme ça?	إذا قلت نعم, ماذا رأيك في الناس الذين يكتبون الدارجة بهذا الشكل؟
Quelle(s) langue(s) vous utilisez quand vous envoyez des SMS à vos amis?	اي لغة او لغات تستخدم(ي) عندما ترسل(ي) الرسائل القصيرة الى أصدقائك بالهاتف؟
Est-ce qu'il est possible d'utiliser l'alphabet arabe sur votre téléphone portable?	هل في هاتفك الحروف العربية؟
Est-ce que vous entendez les gens qui utilisent plus qu'une langue dans une seule conversation parlée?	هل تسمع(ي) أحيانا الناس الذين يستخدمون أكثر من لغة في حوار واحد؟

Si oui, dans quelles situations?	إذا قلت نعم, في اي سياق؟
Quel(s) type(s) de personnes utilisent deux ou plusieurs langues comme ça?	أي نوع من الناس يستخدمون عدة لغات بهذه الطريقة ؟
Selon vous, pourquoi est-ce que les gens utiliseraient les langues comme ça?	في رأيك لماذا يستخدم الناس اللغات بهذا الشكل؟
Est-ce que vous trouvez qu'il existe des situations dans lesquelles il vaut mieux d'utiliser plus qu'une langue?	هل تعتقد أن هناك مواقف من الأحسن استخدام أكثر من لغة واحدة فيها؟

Auto-évaluation de capacité de langue						مستوى اللغة
Directions: Indiquez votre confort dans la langue indiquée quand vous êtes dans les situations données. Utilisez l'échelle de 1 à 5 où 1 = je ne peux pas le faire et 5 = je peux le faire sans problème.						حدد(ي) إذا كنت تستطيع(ين) القيام بالنشاطات التالية باللغة المُحددة من 1 (لا أقدر ان أفعل هذا) حتى 5 (أقدر ان أفعل هذا بدون مشكل)
1 = Je ne peux pas le faire 1 = لا أقدر ان أفعل هذا					5 = Je peux le faire sans problème 5 = أقدر ان أفعل هذا بدون مشكل	
FRANÇAIS	1	2	3	4	5	فرنسية
Commander de la nourriture à un restaurant						طلب في مطعم
Demander le chemin						طلب المساعدة لمعرفة الطريق في المدينة
Acheter les vêtements dans un magasin						شراء الملابس في محل
Expliquer ma santé à un médecin						شرح حالتي الصحية للطبيب
Parler de mes passe-temps						التكلم عن هوايتي
Décrire mes études						التحدث عن دراستي
Parler de mes projets pour l'avenir						التكلم عن خططي المستقبلية
Donner mes avis sur les polémiques						أبين رأيي في المواضيع المهمة
Parler au sujet du gouvernement						التكلم عن الحكومة
Comprendre les conversations quotidiennes						فهم الحوارات العادية
Comprendre les proverbes et les expressions						فهم الامثلة والعبارات
Comprendre les présentations académiques						فهم التقديمات الاكاديمية
Comprendre les blagues et la satire dans les films						فهم النكت و السخرية في الافلام
Comprendre les messages indirects et les allusions						فهم الأفكار غير المباشرة

Auto-évaluation de capacité de langue						مستوى اللغة
Directions: Indiquez votre confort dans la langue indiquée quand vous êtes dans les situations données. Utilisez l'échelle de 1 à 5 où 1 = je ne peux pas le faire et 5 = je peux le faire sans problème.						حدد(ي) إذا كنت تستطيع(ين) القيام بالنشاطات التالية باللغة المُحددة من 1 (لا أقدر ان أفعل هذا) حتى 5 (أقدر ان أفعل هذا بدون مشكل)
1 = Je ne peux pas le faire 1 = لا أقدر ان أفعل هذا					5 = Je peux le faire sans problème 5 = أقدر ان أفعل هذا بدون مشكل	
ARABE STANDARD	1	2	3	4	5	العربية الفصحى
Commander de la nourriture à un restaurant						طلب في مطعم
Demander le chemin						طلب المساعدة لمعرفة الطريق في المدينة
Acheter les vêtements dans un magasin						شراء الملابس في محل
Expliquer ma santé à un médecin						شرح حالتي الصحية للطبيب
Parler de mes passe-temps						التكلم عن هوايتي
Décrire mes études						التحدث عن دراستي
Parler de mes projets pour l'avenir						التكلم عن خططي المستقبلية
Donner mes avis sur les polémiques						أبين رأيي في المواضيع المهمة
Parler au sujet du gouvernement						التكلم عن الحكومة
Comprendre les conversations quotidiennes						فهم الحوارات العادية
Comprendre les proverbes et les expressions						فهم الامثلة والعبارات
Comprendre les présentations académiques						فهم التقديمات الاكاديمية
Comprendre les blagues et la satire dans les films						فهم النكت و السخرية في الافلام
Comprendre les messages indirects et les allusions						فهم الأفكار غير المباشرة

D'AUTRES LANGUES?	لغات أخرى؟
Est-ce que vous parlez d'autres langues? Si oui, lesquelles? Depuis quand ?	هل تتكلم(ي) لغات أخرى؟ إذا قلت نعم, أي لغة أو لغات؟ منذ اي سنة؟

L'Usage de langue idéale					**استخدام اللغات بالطريقة الاحسن**
Veuillez entourer la/les langue(s) que vous trouvez correspond(ent) selon les phrases suivantes:					اختر(ي) اللغة أو اللغات الأحسن في رأيك في كل سياق:
Dans la vie quotidienne, la/les langue(s) la/les plus pratique(s) est/sont:					في الحالة العادية, اللغة أو اللغات المفيدة أكثر:
Le français	mélange de l'arabe marocain et le français	Autre: أخرى:	la langue amazigh	l'arabe marocain	l'arabe standard
فرنسية	خليط دارجة وفرنسية	_____	أمازيغي	دارجة مغربية	فصحى
Je préfère:					الأفضل بالنسبة لي:
Le français	mélange de l'arabe marocain et le français	Autre: أخرى:	la langue amazigh	l'arabe marocain	l'arabe standard
فرنسية	خليط دارجة وفرنسية	_____	أمازيغي	دارجة مغربية	فصحى
La/les langues qui décrit (/décrivent) la culture marocaine le mieux est/sont:					اللغة أو اللغات التي تبين الثقافة المغربية:
Le français	mélange de l'arabe marocain et le français	Autre: أخرى:	la langue amazigh	l'arabe marocain	l'arabe standard
فرنسية	خليط دارجة وفرنسية	_____	أمازيغي	دارجة مغربية	فصحى
Pour être marocain, il faut parler:					لتكون مغربي من الضروري تتكلم:
Le français	mélange de l'arabe marocain et le français	Autre: أخرى:	la langue amazigh	l'arabe marocain	l'arabe standard
فرنسية	خليط دارجة وفرنسية	_____	أمازيغي	دارجة مغربية	فصحى

Appendix B: Shortened Oxford French Placement Test

Test de français Numéro identifiant (de la recherche) _____

Répondez aux questions suivantes. Donnez une réponse à chaque question. Ce test durera entre 5 et 10 minutes. Merci de votre participation.

1. _____ vous parlez français?

a) Qu'est-ce que

b) Est-ce que

c) Quel

d) Quoi

2. _____ 18 ans.

a) Je suis

b) J'ai

c) Je

d) Je suis âgé

3. Monsieur Martin et _____ femme sont très sympathiques.

a) sa

b) son

c) ses

d) leur

4. Vous buvez _____ café.

a) un peu

b) la

c) des

d) du

5. Pour aller à la poste, vous tournez _____.

a) tout droit

b) à la droite

c) à droite

d) droit

6. Les enfants _____ leurs devoirs à 6 heures.

a) finissez
b) finissons
c) finis
d) finissent

7. Hier, nous _____ avec Monsieur le Maire à midi.

a) déjeunions
b) avions déjeuné
c) avons déjeuné
d) a déjeuné

8. Je regardais la télévision depuis 1 heure quand il _____.

a) arrivait
b) va arriver
c) est arrivé
d) arrivera

9. L'année prochaine j'_____ au Canada.

a) irais
b) irai
c) ira
d) irez

10. Mon numéro de téléphone c'est le soixante-dix-huit, quatre-vingt-un, quarante, quatre-vingt-douze _____.

a) 68-41-40-82
b) 78-41-40-92
c) 78-81-40-92
d) 68-81-14-92

11. Je suis arrivé en Angleterre _____ 10 ans.

a) pour
b) depuis
c) il y a
d) pendant

12. Vous avez téléphoné à madame Lesieur? Non, je _____ téléphonerai demain.

a) lui
b) la
c) leur
d) le

13. Vous allez au cinéma? Oui _____ vais souvent.

a) je le
b) j'en
c) j'y
d) je la

14. Il faut que tu _____ à l'aéroport.

a) vas
b) aller
c) ailles
d) iras

15. J'ai lu le livre _____ tu m'avais parlé.

a) dont
b) que
c) qui
d) où

16. À votre place, _____ mes études.

a) je continuerais
b) je continuerai
c) je continue
d) je vais continuer

17. Vous offrez des fleurs à votre femme? Oui, je _____ offre pour son anniversaire.

a) les en
b) en lui
c) lui en
d) en les

18. Le suspect s'est rendu au commissariat où _____ pendant plusieurs heures.

a) il a interrogé
b) il s'est interrogé
c) il a été interrogé
d) il était interrogé

19. Vous sortirez quand vous _____ vos devoirs!

a) aurez fini
b) finirez
c) avez fini
d) auriez fini

20. Je dois envoyer cette lettre _____, c'est urgent.

a) plus tard
b) dans quelques temps
c) dès que possible
d) quand c'est possible

21. Il fait _____ froid que je préfère rester près de la cheminée.

a) beaucoup
b) trop
c) très
d) tellement

22. _____ sa pauvreté, il est heureux.

a) Pourtant
b) Par contre
c) Bien que
d) Malgré

23. Elles se sont _____ une maison au bord de la mer.

a) acheté
b) achetées
c) achetée
d) acheter

24. Je n'ai jamais vu un _____ désordre!

a) similaire

b) tel

c) aussi

d) autant

25. Je ne crois pas que vous _____ raison.

a) avez

b) aurez

c) ayez

d) auriez

26. _____ il pleut, prenons l'autobus.

a) Parce qu'

b) Puisqu'

c) À cause

d) À cause d'

References

Aabi, M. (1999). *The Syntax of Moroccan Arabic/French and Moroccan Arabic/Standard Arabic code-switching*.

Abassi, A. (1977). *A Socio-lingustic analysis of multilingualism in Morocco*. Austin, Texas.

Aboelezz, M. (2009). Latinised Arabic and connections to bilingual ability (pp. 1–23). Retrieved from http://www.lancs.ac.uk/fss/linguistics/pgconference/v03/Volume 03.pdf#page=3

Abu Elhij'a, D. (2012). Facebook Written Levantine Vernacular Languages. *The Levantine Review*, *1*(1), 68–105.

Abu Elhija, D. (2014). A new writing system? Developing orthographies for writing Arabic dialects in electronic media. *Writing Systems Research*, *6*(2), 190–214.

Agence Nationale de Réglementation des Télécommunications (ANRT). (2001). *Rapport Annuel 2001*. Rabat, Morocco.

Agence Nationale de Réglementation des Télécommunications (ANRT). (2014). *Tableau de bord: Marché Internet au Maroc*. Rabat, Morocco.

Aguadé, J. (2010). On vocalism in Moroccan Arabic dialects. In J. P. Monferrer-Sala & N. Al Jallad (Eds.), *The Arabic language across the ages* (pp. 95–105). Wiesbaden: Reichert Verlag.

Aguadé, J., & Benyahia, L. (2013). *Diccionario árabe marroquí: Árabe marroquí español/español árabe-marroquí* (p. 316). Quorum.

Almeman, K., & Lee, M. (2013). Automatic building of Arabic multi dialect text corpora by bootstrapping dialect words. *2013 1st International Conference on Communications, Signal Processing, and Their Applications (ICCSPA)*, 1–6.

Anderson, T. (2006). *Spanish-English bilinguals' attitudes toward code-switching: Proficiency, grammaticality and familiarity*.

Androutsopoulos, J. (2009). "Greeklish": Transliteration practice and discourse in the context of computer-mediated digraphia. In A. Georgakopolou & M. Silk (Eds.), *Standard Languages and Language Standards: Greek, Past and Present* (pp. 221–249). Hampshire, UK: Ashgate Publishing.

Androutsopoulos, J. (2013). Code-switching in computer-mediated communication. *Pragmatics of Computer-Mediated Communication*, 667–694.

Angermeyer, P. S. (2012). Bilingualism meets digraphia: Script alternation and script hybridity in Russian–American writing and beyond. In M. Sebba, S. Mahootian, & C. Jonsson (Eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse* (pp. 255–272). New York: Routledge.

Asharq Al-Awsat. (2012, December 30). Morocco's Facebook users prefer French. *Asharq Al-Awsat*. Retrieved from http://www.aawsat.net/2012/12/article55239322

Atifi, H. (2003). Cultural variation in on-line communications; An Ethnographic study of Moroccan discussion forums. *Language et Société*, *104*(June), 57–82.

Auer, P. (1984). *Bilingual conversation*. John Benjamins Publishing.

Auer, P. (1995). The pragmatics of code-switching: a sequential approach. In L. Milroy & P. M. Muysken (Eds.), *One speaker, two languages: Cross-disciplinary perspectives on code-switching* (pp. 115–135). Cambridge University Press.

Barton, D., & Hamilton, M. (2000). Literacy practices. In D. Barton, M. Hamilton, & R. Ivanič (Eds.), *Situated literacies* (pp. 7–15). London: Routledge.

Bassiouney, R. (2009). *Arabic Sociolinguistics*. Edinburgh University Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package. Retrieved from http://cran.r-project.org/package=lme4

Belazi, H. M. (1991). *Multilingualism in Tunisia and French/Arabic code-switching among educated Tunisian bilinguals*.

Belazi, H. M., Rubin, E. J., & Toribio, A. J. (1994). Code Switching and X-Bar Theory: The Functional Head Constraint. *Linguistic Inquiry*, *25*(2), 221–237.

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204.

Benrabah, M. (2007). Language-in-Education Planning in Algeria: Historical Development and Current Issues. *Language Policy*, *6*(2), 225–252.

Bentahila, A. (1983). *Language attitudes among Arabic-French bilinguals in Morocco*. Multilingual Matters.

Bentahila, A. (1983). Motivations for code-switching among Arabic-French bilinguals in Morocco. *Language & Communication*, *3*(3), 233–243.

Bentahila, A., & Davies, E. E. (1983). The syntax of Arabic-French codeswitching. *Lingua*, *59*, 301–330.

Bentahila, A., & Davies, E. E. (1995). Patterns of code-switching and patterns of language contact. *Lingua*, *96*(2), 75–93.

Bentahila, A., & Davies, E. E. (2002). Language mixing in rai music: localisation or globalisation? *Language & Communication*, *22*(2), 187–207.

Benzakour, F., Gaadi, D., & Queffélec, A. (2000). *Le français au Maroc: Lexique et contacts de langues*. Brussels: Editions Duculot.

Berjaoui, N. (2001). Aspects of the Arabic orthography with preliminary insights from the Moroccan Computer-Mediated Communication. In M. Beißwenger (Ed.), *Chat-Kommunikation: Sprache, Interaction, Sozialität & Indentität in synchroner computerver-mittelter Kommunikation* (pp. 431–435). Stuttgart: Ibidem.

Bianchi, R. M. (2012). 3arabizi - When local Arabic meets global English on the internet. *Acta Linguistica Asiatica*, *2*(1), 89–100.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Bjornsson, J. A. (2010). *Egyptian Romanized Arabic: A Study of Selected Features from Communication Among Egyptian Youth on Facebook*. Oslo, Norway.

Blom, J.-P., & Gumperz, J. J. (1972). Social meaning in linguistic structures: Code-switching in Norway. In J. J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics: The Ethnography of communication* (pp. 407–434). New York: Holt, Rinehart & Winston.

Boumans, L. (1998). *The Syntax of Codeswitching: Analysing Moroccan Arabic / Dutch Conversations*. The Catholic University of Nijmegen.

Bourdieu, P. (1982). *Ce que parler veux dire: l'économie des échanges linguistiques*. Paris: Fayard.

Boutieri, C. (2010). *An Ambivalent embrace: The cultural politics of Arabization and the knowledge economy in the Moroccan public school*. Princeton University.

Brugman, H., & Russel, A. (2004). Annotating Multimedia/ Multi-modal resources with ELAN. *Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation*.

Brustad, K. E. (2000). *The Syntax of spoken Arabic: A comparative study of Moroccan, Egyptian, Syrian and Kuwaiti dialects*. Washington, D.C.: Georgetown University Press.

Bullock, B., & Toribio, A. J. (2009). Themes in the study of code-switching. In *The Cambridge Handbook of Linguistic Code-Switching* (pp. 1–17). Leiden: Cambridge University Press.

Caubet, D. (1983). *La Détermination en Arabe Marocain*. Paris: Université Paris 7.

Caubet, D. (1993). *L'arabe marocain. Tome 1: phonologie et morphosyntaxe. Tome 2: syntaxe et catégories grammaticales, textes.* Paris, Louvain: Peeters.

Caubet, D. (2008). From "Movida" to "nayda" in Morocco: the use of darija (Moroccan Arabic), in artistic creation at the beginning of the 3rd millenium. In S. Prochazka & V. Ritt-Benmimoun (Eds.), *Between the Atlantic and Indian Oceans : studies in contemporary Arabic Dialects* (Neue Beihe., pp. 113–124). Wien: LIT.

Chafik, M. (1999). *al-Darijah al-Maghribiyah: majal tawarud bayna al-Amazighiyah wa-al-`Arabiyah*. Rabat: Akadimiyat al-Mamlakah al-Maghribiyah.

Chakrani, B. (2010). *A Sociolinguistic investigation of language attitudes among youth in Morocco*. University of Illinois at Urbana-Champaign, Urbana-Champaign.

Chakrani, B. (2011). Covert Language Attitudes : A New Outlook on the Sociolinguistic Space of Morocco. In E. G. Bokamba, R. K. Shosted, & B. Tesfaw Ayalew (Eds.), *Selected Proceedings of the 40th Annual Conference on African Linguistics* (pp. 168–177). Somerville, MA: Cascadilla Proceedings Project.

Chakrani, B., & Huang, J. L. (2012). The work of ideology: examining class, language use, and attitudes among Moroccan university students. *International Journal of Bilingual Education and Bilingualism*, *17*(1), 1–14.

Cotterell, R., Renduchintala, A., Saphra, N., & Callison-burch, C. (2011). An Algerian Arabic-French Code-Switched Corpus, 2011–2014.

Crystal, D. (1997). *English as a global language*. Cambridge: Cambridge University Press.

Crystal, D. (2001). *Language and the Internet* (1st ed.). Cambridge, UK; New York: Cambridge University Press.

Crystal, D. (2006). *Language and the internet* (2nd ed.). Cambridge, UK; New York: Cambridge University Press.

Daoud, M. (2011). The sociolinguistic situation in Tunisia: language rivalry or accommodation? *International Journal of the Sociology of Language*, *2011*(211).

Davies, E. E., & Bentahila, A. (1989). On mother and other tongues: The Notion of Possession of a Language. *Lingua*, *78*, 267–293.

Deuber, D., & Hinrichs, L. (2007). Dynamics of orthographic standardization in Jamaican Creole and Nigerian Pidgin. *World Englishes*, *26*(1), 22–47.

Deuchar, M., Muysken, P., & Wang, S.-L. (2007). Structured variation in codeswitching: towards an empirically based typology of bilingual speech patterns. *International Journal of Bilingual Education and Bilingualism*, *10*(3), 298–340.

Deumert, A., & Masinyana, O. (2008). Mobile language choices – the use of English and isiXhosa in text messages (SMS): Evidence from a bilingual South African sample. *English World-Wide*, *29*, 117–148.

Diab, M., Habash, N., Rambow, O., Altantawy, M., & Benajiba, Y. (2010). COLABA: Arabic dialect annotation and processing (pp. 66–74). Retrieved from http://www.lrec-conf.org/proceedings/lrec2010/workshops/W1.pdf#page=71

DiSciullo, A.-M., Muysken, P., & Singh, R. (1986). Government and Code-Mixing. *Journal of Linguistics*, *22*(1), 1–24.

Dorleijn, M., & Nortier, J. (2009). Code-switching and the internet. In B. Bullock & A. J. Toribio (Eds.), *The Cambridge Handbook of Linguistic Code-Switching* (pp. 127–141). Leiden: Cambridge University Press.

Eckert, P. (1989). *Jocks and burnouts: Social categories and identity in the high school*. Teachers College Press.

Eid, M. (2002). Language is a choice: Variation in Egyptian Women's written discourse. In A. Rouchdy (Ed.), *Language contact and language conflict in Arabic: Variations on a sociolinguistic theme* (pp. 203–232). London: Routledge.

Ennaji, M. (2002). Language contact, arabization policy and education in Morocco. In A. Rouchdy (Ed.), *Language contact and language conflict in Arabic: Variations on a sociolinguistic theme* (pp. 70–90). London: Routledge.

Ennaji, M. (2005). *Multilingualism, Cultural Identity, and Education in Morocco*. Springer.

Ennaji, M., & Sadiqi, F. (1992). *Introduction to modern linguistics*. Casablanca: Afrique-Orient.

Espinosa, A. M. (1911). Studies in New Mexican Spanish, Part II: Morphology. *Revue de Dialectologie Romane*, *3*, 251–286.

Federation national du tourisme. (2010). *10eme Assises du tourisme*. Rabat, Morocco.

Ferguson, C. A. (1959). Diglossia. *Word - Journal of the International Lingusitic Association*, *15*(2), 325–340.

Fishman, J. A. (1967). Bilingualism with and without diglossia; Diglossia with and without bilingualism. *Journal of Social Issues*, *23*(2), 29–38.

FL: Wikimedia Foundation Inc. (2015). Wikipedia: The free encyclopedia. Retrieved February 23, 2015 from https://arz.wikipedia.org/wiki/الرئيسيه_الصفحه

Gardner-Chloros, P. (2009). *Code-switching*. Cambridge: Cambridge University Press.

Gee, J. P. (1996). *Social linguistics and literacies: Ideology in Discourses* (Second Edition). London: Taylor & Francis.

Grabowski, J. (2011). *Stability of Grammaticality Judgments in German-English Code-Switching*. Unpublised Master's Thesis. Arizona State University. Retrieved from http://repository.asu.edu/attachments/56461/content/Grabowski_asu_0010N_10427.pdf

Grover, D. L., Camp, J. L. J., & Belnap, N. D. J. (1975). A Prosentential theory of truth. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *27*(2), 73–125.

Gullberg, M., Indefrey, P., & Muysken, P. (2009). Research techniques for the study of code-switching. In B. E. Bullock & A. J. Toribio (Eds.), *The Cambridge Handbook of Linguistic Code-Switching* (pp. 21–39). Cambridge University Press.

Habib, R. (2010a). Towards determining social class in Arabic-speaking communities and implications for linguistic variation. *Sociolinguistic Studies*, *4*(1), 175–200.

Habib, R. (2010b). Towards determining social class in Arabic-speaking communities and implications for linguistic variation. *Sociolinguistic Studies*, *4*(1), 175–2000.

Hachimi, A. (2007). Becoming Casablancan. In C. Miller, E. Al-Wer, D. Caubet, & J. C. E. Watson (Eds.), *Arabic in the city: Issues in dialect contact and language variation*. Routledge.

Hachimi, A. (2013). The Maghreb-Mashreq language ideology and the politics of identity in a globalized Arab world 1, 269–296.

Harrell, R. S. (1962). *A short reference grammar of Moroccan Arabic*. Washington, D.C.: Georgetown University Press.

Harrell, Richard Slade (Ed.). (1966). *A dictionary of Moroccan Arabic: Moroccan-English, English-Moroccan.* Washington, D.C.: Georgetown Univ Pr.

Harris, J. (2010). *The frequency and distribution of written and spoken anglicisms in two varieties of French*. Concordia University, Montreal, Quebec.

Haut-Commissariat au Plan. (2004). *Recensement Général de la Population et de l'Habitat*.

Heath, J. (1989). *From code-switching to borrowing: foreign and diglossic mixing in Moroccan Arabic.* London: Kegan Paul International.

Heath, J. (1997). Moroccan Arabic Phonology. In *Phonologies of Asia and Africa (including the Caucasus)* (pp. 205–217). Eisenbrauns.

Heath, J. (2002). *Jewish and Muslim dialects of Moroccan Arabic*. Routledge.

Herring, S. C. (2001). Computer-mediated Discourse. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds.), *The Handbook of Discourse Analysis* (pp. 612–634). Oxford: Blackwell.

Herring, S. C. (2007). A Faceted Classification Scheme for Computer - Mediated Discourse. *Language@Internet*, *1*, 1–37. Retrieved from http://www.languageatinternet.de/articles/2007/761/index_html

Herring, S. C. (2010). Computer-mediated conversation: Introduction and overview. *Language@Internet*, *8*(2).

Herring, S. C. (2013). Discourse in Web2.0: Familiar, Reconfigured, and Emergent. In *Georgetown University Round Table on Languages and Linguistics 2011: Discourse 2.0: Language and new media* (pp. 1–29). Washington, D.C.: Georgetown University Press.

High Commission for Planning. (2009). *At-Tabaqaat al-wusta Al-Maghribiyya: Khasaasuhaa, Tattawuruhaa, wa Awaamil Tawusuuha, Al-Juza Ath-Thaani*. Al-Mamlaka Al-Maghribiyya Al-Mandubiyya As-Saamiyya Li Ttakhtiit.

Hinrichs, L. (2004). Emerging orthographic conventions in written creole: Computer-mediated communication in Jamaica. *Arbeiten Aus Anglistik Und Amerikanistik, Band 29*, 81–109.

Hinrichs, L. (2006). *Codeswitching on the web: English and Jamaican Creole in e-mail communication* (Vol. 147.). John Benjamins Publishing.

Hinrichs, L., & Deuber, D. (2007). Dynamics of orthographic standardization in Jamaican Creole and Nigerian Pidgin. *World Englishes*, *26*(1), 22–47.

Hinrichs, L., & White-Sustaíta, J. (2011). Global Englishes and the sociolinguistics of spelling: A study of Jamaican blog and email writing. *English World-Wide*, *32*(1), 46–73.

Ibrahim, M. H. (1986). Standard and Prestige Language: A Problem in Arabic Sociolinguistics. *Anthropological Linguistics*, *28*(1), 115–126.

Joshi, A. (1985). Processing of sentences with intrasentential code switching. In et al. Dowty (Ed.), *Natural Language Parsing: Psychological, computational and theoretical persepectives*. Cambridge: Cambridge University Press.

Knibiehler, Y., Emmery, G., & Leguay, F. (1992). *Des Français au Maroc: La presence et la mémoire*. Paris: Editions Denoël.

Kossman, M. (2012). Berber. In Z. Frajzynigier & E. Shay (Eds.), *The Afroasiatic Languages*. Cambridge University Press.

Kytölä, S. (2012). Multilingual web discussion forums: Theoretical, practical and methodological issues. In M. Sebba, S. Mahootian, & C. Jonsson (Eds.), *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse* (pp. 106–127). New York: Routledge.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics, Washington.

Labov, W. (1991). The intersection of sex and social class in the course of linguistic change, *2*(1990), 205–254.

Labov, W. (2001). *Principles of Lingusitic Change: Social Factors*. Malden, Oxford: Wiley-Blackwell.

Lahlou, M. (1991). *A morpho-syntactic study of code-switching between Moroccan Arabic and French*. The University of Texas at Austin.

Lawson, S., & Sachdev, I. (2000). Codeswitching in Tunisia: Attitudinal and behavioural dimensions. *Journal of Pragmatics*, *32*(9), 1343–1361.

Lawson-Sako, S., & Sachdev, I. (1997). Linguistic differentiation in Tunisia: Status and solidarity dimensions. In H. Ajroud (Ed.), *The Canon* (pp. 175–196). Tunis: Faculté de Lettres de la Manouba.

Lee, C. K.-M. (2007). Affordances and Text-Making Practices in Online Instant Messaging. *Written Communication*, *24*(3), 223–249.

Lexander, K. V. (2012). Analyzing multilingual texting in Senegal: An Approach for the study of mixed-language SMS. In M. Sebba, S. Mahootian, & C. Jonsson (Eds.), *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse* (pp. 146–169). New York: Routledge.

Maas, U. (2011). *Marokkanisches Arabisch. Die Grundstrukturen*. LINCOM publishers.

Mackey, W. F. (1970). Interference, integration and the synchronic fallacy. Laval University, Quebec, International Center on Bilingualism.

MacSwan, J. (1997). *A minimalist approach to intrasentential code switching: Spanish-Nahuatl bilingualism in Central Mexico*.

MacSwan, J. (2009). Generative approaches to code-switching. In B. Bullock & A. J. Toribio (Eds.), *The Cambridge Handbook of Linguistic Code-Switching* (pp. 309–335). Leiden: Cambridge University Press.

MacSwan, J. (2014). Programs and proposals in codeswitching research: unconstraining theories of bilingual language mixing. In J. MacSwan (Ed.), *Grammatical Theory and Bilingual Codeswitching* (pp. 1–33). MIT Press.

Magidow, A. (2013). Towards a Sociohistorical Reconstruction of Pre-Islamic Arabic.

Mahootian, S., & Santorini, B. (1996). Code Switching and the Complement/Adjunct Distinction. *Linguistic Inquiry*, *27*(3), 464–479. Retrieved from

Marçais, P. (1977). *Esquisse grammaticale de l'arabe maghrébin*. Librairie d'Amérique et d'Orient.

Maroc Service des Statistiques. (1960). *Denombrement générale de la population de la zone française de l'Empire Cherifien effectué 1960.* Rabat, Morocco.

Mercier, H. (1959). *Dictionnaire français-arabe* (2nd ed.). Rabat, Tanger: Les éditions La Porte, Les éditions Eurafrique.

Micaud, C. A. (1974). Bilingualism in North Africa: cultural and sociopolitical implications. *The Western Political Quarterly*, 92–103.

Milroy, J. (1981). *Regional accents of English: Belfast*. Belfast: Blackstaff.

Montes-Alcalá, C. (2000). Written Code-switching: Powerful Bilingual Images. In R. Jacobson (Ed.), *Codeswitching Worldwide II* (pp. 59–74). Berlin: Mouton.

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge University Press.

Myers-Scotton, C. (1992). Comparing codeswitching and borrowing. *Journal of Multilingual & Multicultural Development*, *13*(1-2), 19–39.

Myers-Scotton, C. (1993). *Duelling languages; Grammatical structure in code-switching*. Oxford: Oxford University Press.

Myers-Scotton, C. (2002). *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes.* Oxford: Oxford University Press.

Myers-Scotton, C. (2008). Language contact: Why outsider system morphemes resist transfer. *Journal of Language Contact*, *2*(1), 21–41.

Myers-Scotton, C., & Jake, J. L. (2000). Four types of morpheme: Evidence from aphasia, code switching, and second-language acquisition. *Linguistics*, *38*, 1053–1100.

Myers-Scotton, C., & Jake, J. L. (2009). A universal model of code-switching and bilingual language processing and production. In B. E. Bullock & A. J. Toribio (Eds.), *The Cambridge Handbook of Linguistic Code-Switching* (pp. 336–357). Cambridge University Press.

Nait M'Barek, M., & Sankoff, D. (1988). Le discours mixte arabe/français: emprunts ou alternances de langue? *Revue Canadienne de Linguistique*, *33*(2), 143–154.

Naylor, P. C. (2009). *North Africa: A History from antiquity to the present*. Austin, TX: University of Texas Press.

Nortier, J. M. (1989). *Dutch and Moroccan Arabic in contact: Code switching among Moroccans in the Netherlands*. University of Amsterdam.

Nortier, J. M. (1995). Code switching in Moroccan Arabic/Dutch vs. Moroccan Arabic/French language contact. *International Journal of the Sociology of Language*, *112*(1), 81–95.

Nurmi, A., & Pahta, P. (2012). Multilingual practices in women's English correspondence 1400-1800. In M. Sebba, S. Mahootian, & C. Jonsson (Eds.), *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse* (pp. 44–67). New York: Routledge.

Otheguy, R., & Lapidus, N. (2003). An adaptive approach to noun gender in New York contact Spanish. In *Amsterdam Studies in the theory and history of Linguistic Science series 4* (pp. 209–232).

Oukarfi, S. (2013). L'usage de l'Internet au Maroc : Essai de mesure de la fracture numérique de second degré [The use of the Internet in Morocco : Test for measuring the second digital divide], *2*(2), 118–130.

Oxford University Language Center. (n.d.). Oxford Placement Test. Retrieved from http://www.lang.ox.ac.uk/courses/tst_placement_french.html

Palfreyman, D., & Al Khalil, M. (2003). "A Funky Language for Teenzz to Use:" Representing Gulf Arabic in Instant Messaging. *Journal of Computer-Mediated Communication*, *9*(1).

Pfaff, C. W. (1979). Constraints on Language Mixing: Intrasentential Code-Switching and Borrowing in Spanish/English. *Language*, *55*(2), 291–318.

Poplack, S. (1980a). Sometimes i'll start a sentence in Spanish y termino en espanol: toward a typology of code-switching. *Linguistics*, *18*(7-8), 581–618.

Poplack, S. (1980b). Sometimes ΙΙΙ start a sentence in Spanish Y TERMINO EN ESPANOL : toward a typology of code-switching. *Linguistics*, *18*(1980), 581–618.

Poplack, S. (1981). Syntactic Structure and social function of Codeswitching. In R. P. Duran (Ed.), *Latino Language and Communicative Behavior* (pp. 169–184). Norwood, New Jersey: ABLEX Publishing Corporation.

Poplack, S., & Meechan, M. (1988). Patterns of language mixture: Nominal structure in Wolof-French and Fonge-French bilingual discourse. In L. Milroy & P. Muysken (Eds.), *One Speaker, Two Languages.* Cambridge: Cambridge University Press.

Poplack, S., & Sankoff, D. (1984). Borrowing: the synchrony of integration. *Linguistics*, *22*(1984), 99–135.

Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, *26*(1), 47–104.

Post, R. (2010). French in written Arabic-French code-switching among young Moroccan Arabic Speakers Rate of French use Parts of Speech in French.

Post, R. E. (2010). *Code-switching in the Determiner Phrase : A comparison of Tunisian Arabic-French and Moroccan Arabic-French switching*. The University of Texas at Aystub.

R Core Team. (2014). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org

Redouane, R. (2005). Linguistic constraints on codeswitching and codemixing of bilingual Moroccan Arabic-French speakers in Canada. In J. Cohen, K. T. McAlister, K. Roland, & J. MacSwan (Eds.), *ISB4: Proceedings of the 4th International Symposium on Bilingualism* (pp. 1921–1933). Somerville, MA: Cascadilla Proceedings Project.

Romaine, S. (1989). *Bilingualism*. (Basil Blackwell, Ed.). Oxford.

Sankoff, D., & Poplack, S. (1981). A formal grammar for code-switching. *Research on Language & Social Interaction*, *14*(1), 3–45.

Sayahi, L. (2011a). Code-switching and language change in Tunisia. *International Journal of the Sociology of Language*, *2011*(211).

Sayahi, L. (2011b). Spanish in Contact with Arabic. In M. Diaz-Campos (Ed.), *The Handbook of Hispanic Sociolinguistics* (pp. 473–489). Wiley-Blackwell.

Sayahi, L. (2014). *Diglossia and language contact: Language variation and change in North Africa*. Cambridge: Cambridge University Press.

Schendl, H. (2012). Literacy, Multilingualism and Code-switching in early English writing. In M. Sebba, S. Mahootian, & C. Jonsson (Eds.), *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse* (pp. 27–43). New York: Routledge.

Scipione, R., & Sayahi, L. (2005). Consonantal Variation of Spanish in Northern Morocco. In L. Sayahi & M. Westmoreland (Eds.), *Selected Proceedings of the second workshop on Spanish Sociolinguistics* (pp. 127–132). Somerville, MA: Cascadilla Proceedings Project.

Sebba, M. (1998a). A congruence approach to the syntax of codeswitching. *International Journal of Bilingualism*, *2*(1), 1–19. Retrieved from http://ijb.sagepub.com/content/2/1/1.short

Sebba, M. (1998b). Phonology meets ideology: The meaning of orthographic practices in British Creole. *Language Problems & Language Planning*, *22*(1), 19–47.

Sebba, M. (2007). *Spelling and society: The culture and politics of orthography around the world*. Cambridge University Press.

Sebba, M. (2012). Writing Switching in British Creole. In M. Sebba, S. Mahootian, & C. Jonsson (Eds.), *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse* (pp. 89–105). New York: Routledge.

Sebba, M., Mahootian, S., & Jonsson, C. (2012). *Language Mixing and Code-Switching in Writing*.

Sirles, C. A. (1999). Politics and Arabization: the evolution of postindependence North Africa. *International Journal of the Sociology of Language*, *137*(1), 115–130.

Stevens, P. B. (1974). *French and Arabic bilingualism in North Africa with special reference to Tunisia: a study of attitudes and language use patterns.* Georrgetown University.

Stroomer, H. (2008). Three Tashelhiyt Berber texts from the Arsène Roux archives (Vol. 2, pp. 389–397).

Tagliamonte, S. A., & Denis, D. (2008). Linguistic Ruin? Lol! Instant Messaging and Teen Language. *American Speech*, *83*(1), 3–34.

Themistocleous, C. (2010). Writing in a non-standard Greek variety: Romanized Cypriot Greek in online chat. *Writing Systems Research, 2*(2), 155–168.

Timm, L. A. (1975). Spanish-English Code-switching: El porque y how-not-to. *Romance Philology, 284*, 473–482.

Treffers-Daller, J. (1994). *Mixing two languages: French-Dutch contact in a comparative perspective*. Berlin: Mouton de Gruyter.

Treffers-Daller, J. (2009). *Code-switching and transfer: an exploration of similarities and differences*. Cambridge: Cambridge University Press.

Turner, M. L. (2013). *Definiteness marking in Moroccan Arabic: contact, divergence, and semantic change*. The University of Texas at Austin, Austin, Texas. Retrieved from https://repositories.lib.utexas.edu/handle/2152/21189

Valdès-Fallis, G. (1977). The sociolinguistics of Chicano literature: Towards an analysis of the role and function of language alternation in contemporary bilingual poetry. *Point of Contact/Punto de Contacto, 1*(4), 30–39.

Van Hout, R., & Muysken, P. (1994). Modeling lexical borrowability. *Language Variation and Change*, *6*(1), 39–62.

Vicente, A., & Ziamari, K. (2008). L'arabe marocain au contact du français et de l'espagnol. In S. Procházka & V. Ritt-Benmimoun (Eds.), *Between the Atlantic and Indian Oceans : studies in contemporary Arabic Dialects. Proceedings of the 7th AIDA Conference, Vienne, Neue Beihefte zur Wiener Zeitschrift für die Kunde des Morgenlandes Bd* (pp. 457–469).

Wagner, D.A. (1993). *Literacy, Culture and Development: Becoming Literate in Morocco.* New York: Cambridge University Press.

Walters, K. (1998). Fergie's prescience: The changing nature of diglossia in Tunisia. *Int'l. J, 165*(2516/03), 0163–0077.

Warschauer, M., El Said, G. R., & Zohry, A. G. (2006). Language Choice Online: Globalization and Identity in Egypt. *Journal of Computer-Mediated Communication, 7*(2002), 1–18.

Wehr, H. (1994). *Arabic-English Dictionary*. (J. M. Cowan, Ed.) (4th ed.). Urbana, Illinois: Spoken Language Services, Inc.

Weinreich, U. (1953). *Languages in Contact*. The Hague: Mouton.

Wodak, R., & Benke, G. (1997). Gender as a sociolinguistic variable: New perspectives on variation studies. In F. Coulmas (Ed.), *The Handbook of Sociolinguistics* (pp. 127–150). Malden, Oxford, Victoria: Blackwell.

Woolford, E. (1983). Bilingual code-switching and syntactic theory. *Linguistic Inquiry, 14*(3), 520–536.

World Bank. (2009). Data Profile: Morocco. Retrieved from http://data.worldbank.org/country/morocco

Yaghan, M. A. (2008). "Arabizi": A Contemporary Style of Arabic Slang. *Design Issues, 24*(2), 39–52.

Youssi, A. (1992). *Grammaire et Lexique de l'arabe marocain moderne*. Wallada.

Youssi, A. (1995). The Moroccan triglossia: facts and implications. *International Journal of the Sociology of Language, 112.1*, 29–44.

Zenner, E., Speelman, D., & Geeraerts, D. (2012). Cognitive Sociolinguistics meets loanword research: Measuring variation in the success of anglicisms in Dutch. *Cognitive Linguistics, 23*(4), 749–792.

Ziamari, K. (2003). *Le code switching intra-phrastique dans les conversations des étudiants marocains de l'ENSAM: Approche linguistique du duel entre l'arabe marocain et le français.* INALCO (Paris) and Faculty of Letters and Human Sciences (Fez).

Ziamari, K. (2008). *Le code switching au Maroc: l'arabe marocain au contact du français*. Paris: Editions L'Harmattan.

Ziamari, K. (2009). Le contact entre l'arabe marocain et le français au Maroc: spécificités linguistique et sociolinguistique. *Synergies Tunisie, 1*, 173–186. Retrieved from http://ressources-cla.univ-fcomte.fr/gerflint/Tunisie1/ziamari.pdf

Zoabi, Z. (2012). *A'amiya : kef mnektibha ? Alphabet Choice in Electronic A'amiya in Israel and The Arab World*. University of Haifa.